

# LARGE DEVIATIONS FOR MULTISCALE DIFFUSIONS VIA WEAK CONVERGENCE METHODS

PAUL DUPUIS AND KONSTANTINOS SPILIOPOULOS

*Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University,  
Providence, RI 02912*

**ABSTRACT.** We study the large deviations principle for locally periodic stochastic differential equations with small noise and fast oscillating coefficients. There are three possible regimes depending on how fast the intensity of the noise goes to zero relative to the homogenization parameter. We use weak convergence methods which provide convenient representations for the action functional for all three regimes. Along the way we study weak limits of related controlled SDEs with fast oscillating coefficients and derive, in some cases, a control that nearly achieves the large deviations lower bound at the prelimit level. This control is useful for designing efficient importance sampling schemes for multiscale diffusions driven by small noise.

**Keywords:** Large deviations, multiscale diffusions, importance sampling, rugged energy landscape.

## 1. INTRODUCTION

The purpose of this paper is to obtain large deviation properties of stochastic differential equations with rapidly fluctuating coefficients in a form that can be used for accelerated Monte Carlo. Such results are not available in the literature. We use methods from weak convergence and stochastic control. Consider the  $d$ -dimensional process  $X^\epsilon \doteq \{X_t^\epsilon, 0 \leq t \leq 1\}$  satisfying the stochastic differential equation (SDE)

$$(1.1) \quad dX_t^\epsilon = \left[ \frac{\epsilon}{\delta} b\left(X_t^\epsilon, \frac{X_t^\epsilon}{\delta}\right) + c\left(X_t^\epsilon, \frac{X_t^\epsilon}{\delta}\right) \right] dt + \sqrt{\epsilon} \sigma\left(X_t^\epsilon, \frac{X_t^\epsilon}{\delta}\right) dW_t, \quad X_0^\epsilon = x_0,$$

where  $\delta = \delta(\epsilon) \downarrow 0$  as  $\epsilon \downarrow 0$  and  $W_t$  is a standard  $d$ -dimensional Wiener process. The functions  $b(x, y)$ ,  $c(x, y)$  and  $\sigma(x, y)$  are assumed to be smooth according to Condition 2.2 and periodic with period 1 in every direction with respect to the second variable.

If  $\delta$  is of order 1 while  $\epsilon$  tends to zero, large deviations theory tells how quickly (1.1) converges to the deterministic ODE given by setting  $\epsilon$  equal to zero. If  $\epsilon$  is of order 1 while  $\delta$  tends to zero, homogenization occurs and one obtains an equation with homogenized coefficients. If the two parameters go to zero together then one expects different behaviors depending on how fast  $\epsilon$  goes to zero relative to  $\delta$ .

---

*E-mail address:* Paul.Dupuis@brown.edu, kspiliop@dam.brown.edu.

Research of P.D. supported in part by the National Science Foundation (DMS-0706003, DMS-1008331), the Department of Energy (DE-SC0002413), and the Army Research Office (W911NF-09-1-0155).

Research of K.S. supported in part by the Department of Energy (DE-SC0002413).

Using the weak convergence approach of [18], we investigate the large deviations principle (LDP) of  $X^\epsilon$  under the following three regimes:

$$(1.2) \quad \lim_{\epsilon \downarrow 0} \frac{\epsilon}{\delta} = \begin{cases} \infty & \text{Regime 1,} \\ \gamma \in (0, \infty) & \text{Regime 2,} \\ 0 & \text{Regime 3.} \end{cases}$$

The weak convergence approach results in a convenient representation formula for the large deviations action functional (otherwise known as the rate function) for all three regimes (Theorem 2.9). It is based on the representation Theorem 2.4, which in this case involves controlled SDE's with fast oscillating coefficients. Along the way, we obtain a uniform proof of convergence of the underlying controlled SDE (CSDE) in all three regimes (Theorem 2.8). In addition, in some cases we construct a control that nearly achieves the large deviations lower bound at the prelimit level. This control is useful, in particular, for the design of efficient importance sampling schemes. The particular use of the control will appear elsewhere.

A motivation for this work comes from chemical physics and biology, and in particular from the dynamical behavior of proteins such as their folding and binding kinetics. It was suggested long ago (e.g., [34]) that the potential surface of a protein might have a hierarchical structure with potential minima within potential minima. The underlying energy landscapes of certain biomolecules can be rugged (i.e., consist of many minima separated by barriers of varying heights) due to the presence of multiple energy scales associated with the building blocks of proteins. Roughness of the energy landscapes that describe proteins has numerous effects on their folding and binding as well as on their behavior at equilibrium. Often, these phenomena are described mathematically by diffusion in a rough potential where a smooth function is superimposed by a rough function (see Figure 5.1). A representative, but by no means complete, list of references is [6, 16, 24, 35, 39, 42]. The situation investigated in these papers is only a special case of equation (1.1) with  $\sigma(x, y) = \sqrt{2D}$ ,  $b(x, y) = -\frac{2D}{k_\beta T} \nabla Q(y)$  and  $c(x, y) = -\frac{2D}{k_\beta T} \nabla V(y)$ , where  $k_\beta$  is the Boltzmann constant and  $T$  is the temperature. The questions of interest in these papers are related to the effect of taking  $\delta \downarrow 0$  with  $\epsilon$  small but fixed. This is almost the same to requiring that  $\delta$  goes to 0 much faster than  $\epsilon$  does. Our goal is to study the related large deviations principle, so we take  $\epsilon \downarrow 0$  as well. It will become clear that the formula for the effective diffusivity (denoted by  $q$  in Corollary 5.4) that appears in the aforementioned chemistry and biology literature is obtained under Regime 1.

Singularly perturbed stochastic control problems and related large deviations problems have been studied elsewhere (see for example [8, 13, 20, 22, 30, 31, 33, 36, 40, 41] and the references therein). In particular, in [20] the authors study the large deviation problem for periodic coefficients, i.e.,  $b(x, y) = b(y)$ ,  $c(x, y) = c(y)$  and  $\sigma(x, y) = \sigma(y)$ , using other methods. In [20], the authors provide an explicit formula for the action functional in Regime 1, whereas in Regimes 2 and 3 the action functional is in terms of solutions to variational problems. In the present paper, we derive the same explicit expression for the action functional in Regime 1. In addition, we also obtain the related control that nearly achieves the LDP lower bound at the prelimit level. For Regimes 2 and 3 we provide an alternative expression, from [20], for the action functional (Theorem 2.9). It follows from these expressions that Regime 3 can be seen as a limiting case of Regime 2 by simply setting  $\gamma = 0$ , though we are able to prove the large deviation lower bound in Regime 3 only under additional conditions. For both regimes we derive explicit expressions for the action functional in special cases of interest, and in Regime 2 obtain a corresponding control that nearly achieves the LDP lower bound. Note that the extension of the results of [20] for Regime 2 to include the  $x$ -dependence is non-trivial, since several smoothness properties of the local rate function need to be proven (see Subsection 6.1 for details). Apart from [20], Regime 2 has also been studied in [31, 40, 41] under various assumptions and dependencies of the coefficients of the system on the slow and fast motion.

In [20, 40, 41], the local rate function is characterized as the Legendre-Fenchel transform of the limit of the normalized logarithm of an exponential moment or of the first eigenvalue of an associated operator. In the present paper, we provide a direct expression for the local rate function (Theorem 6.1).

We note here that in the case of Regime 1 one can weaken the periodicity assumption, using the results of [36] and the methodology of the present paper, and prove an analogous result when the fast variable takes values in  $\mathbb{R}^d$ . It also seems possible to combine the methods of the present paper together with results in [26, 10] to weaken the periodicity assumption for Regime 2 as well; see Remark 2.12 for more details.

The paper is organized as follows. In Section 2, we establish notation, review some preliminary results and state the general large deviations result (Theorem 2.9). Section 3 considers the weak limit of the associated controlled stochastic differential equations. In Section 4 we prove the large deviations upper bound for all three regimes and the compactness of the level sets of the rate function. Section 5 contains the proof of the large deviations lower bound (or equivalently Laplace principle upper bound) for Regime 1, which completes the proof of the large deviations principle for Regime 1. This section also discusses an explicit expression for a control that nearly achieves the large deviations lower bound in the prelimit level ( $\epsilon > 0$ ). In Section 6, we prove the large deviations lower bound for Regime 2 and identify a control that nearly achieves this lower bound. Section 7 discusses the large deviations lower bound principle for Regime 3 and presents alternative expressions for the rate function in dimension 1.

## 2. PRELIMINARIES, STATEMENT OF THE MAIN RESULTS.

We work with the canonical filtered probability space  $(\Omega, \mathfrak{F}, \mathbb{P})$  equipped with a filtration  $\mathfrak{F}_t$  that satisfies the usual conditions, namely,  $\mathfrak{F}_t$  is right continuous and  $\mathfrak{F}_0$  contains all  $\mathbb{P}$ -negligible sets.

In preparation for stating the main results, we recall the concept of a Laplace principle. Throughout this paper only random variables that take values in a Polish space are considered. By definition, a rate function on a Polish space  $\mathcal{S}$  maps  $\mathcal{S}$  into  $[0, \infty]$  and has compact level sets.

**Definition 2.1.** *Let  $\{X^\epsilon, \epsilon > 0\}$  be a family of random variables taking values in  $\mathcal{S}$  and let  $I$  be a rate function on  $\mathcal{S}$ . We say that  $\{X^\epsilon, \epsilon > 0\}$  satisfies the Laplace principle with rate function  $I$  if for every bounded and continuous function  $h : \mathcal{S} \rightarrow \mathbb{R}$*

$$\lim_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] = \inf_{x \in \mathcal{S}} [I(x) + h(x)].$$

A Laplace principle is equivalent to the corresponding large deviations principle with the same rate function (if the definition of a rate function includes the requirement of compact level sets, see Theorems 2.2.1 and 2.2.3 in [18]). Thus instead of proving a large deviations principle for  $\{X^\epsilon\}$  we prove a Laplace principle for  $\{X^\epsilon\}$ .

Regarding the SDE (1.1) we impose the following condition.

**Condition 2.2.** (i) *The functions  $b(x, y), c(x, y), \sigma(x, y)$  are Lipschitz continuous and bounded in both variables and periodic with period 1 in the second variable in each direction. In the case of Regime 1 we additionally assume that they are  $C^1(\mathbb{R}^d)$  in  $y$  and  $C^2(\mathbb{R}^d)$  in  $x$  with all partial derivatives continuous and globally bounded in  $x$  and  $y$ .*  
(ii) *The diffusion matrix  $\sigma \sigma^T$  is uniformly nondegenerate.*

The regularity conditions imposed are stronger than necessary, but they are assumed to simplify the exposition. See Remark 2.11 for some further details on this. For notational convenience we define the operator  $\cdot : \cdot$ , where for two matrices  $A = [a_{ij}], B = [b_{ij}]$

$$A : B \doteq \sum_{i,j} a_{ij} b_{ij}.$$

Under Regime 1, we also impose the following condition.

**Condition 2.3.** Let  $\mu(dy|x)$  be the unique invariant measure corresponding to the operator

$$\mathcal{L}_x^1 = b(x, y) \cdot \nabla_y + \frac{1}{2} \sigma(x, y) \sigma(x, y)^T : \nabla_y \nabla_y$$

equipped with periodic boundary conditions in  $y$  ( $x$  is being treated as a parameter here). Under Regime 1, we assume the standard centering condition (see [9]) for the unbounded drift term  $b$ :

$$\int_{\mathcal{Y}} b(x, y) \mu(dy|x) = 0,$$

where  $\mathcal{Y} = \mathbb{T}^d$  denotes the  $d$ -dimensional torus.

We note that under Conditions 2.2 and 2.3, for each  $\ell \in \{1, \dots, d\}$  there is a unique, twice differentiable function  $\chi_\ell(x, y)$  that is one periodic in every direction in  $y$ , that solves the following cell problem (for a proof see [9], Theorem 3.3.4):

$$(2.1) \quad \mathcal{L}_x^1 \chi_\ell(x, y) = -b_\ell(x, y), \quad \int_{\mathcal{Y}} \chi_\ell(x, y) \mu(dy|x) = 0.$$

We write  $\chi = (\chi_1, \dots, \chi_d)$ .

Our tool for proving the Laplace principle will be the weak convergence approach of [18]. The following representation theorem is essential for this approach. A proof of this theorem is given in [14]. The control process can depend on  $\epsilon$  but this is not always denoted explicitly. In the representation and elsewhere we take  $T = 1$ . Analogous results hold for arbitrary  $T \in (0, \infty)$ .

**Theorem 2.4.** Assume Condition 2.2, and given  $\epsilon > 0$  let  $X^\epsilon$  be the unique strong solution to (1.1). Then for any bounded Borel measurable function  $h$  mapping  $\mathcal{C}([0, 1]; \mathbb{R}^d)$  into  $\mathbb{R}$

$$-\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] = \inf_{u \in \mathcal{A}} \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|u_t\|^2 dt + h(\bar{X}^\epsilon) \right],$$

where  $\mathcal{A}$  is the set of all  $\mathfrak{F}_t$ -progressively measurable  $d$ -dimensional processes  $u \doteq \{u_t, 0 \leq t \leq 1\}$  satisfying

$$\mathbb{E} \int_0^1 \|u_t\|^2 dt < \infty,$$

and  $\bar{X}^\epsilon$  is the unique strong solution to

$$(2.2) \quad d\bar{X}_t^\epsilon = \left[ \frac{\epsilon}{\delta} b \left( \bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta} \right) + c \left( \bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta} \right) \right] dt + \sigma \left( \bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta} \right) u_t dt + \sqrt{\epsilon} \sigma \left( \bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta} \right) dW_t, \quad \bar{X}_0^\epsilon = x_0.$$

Before stating the main results, we need additional notation and definitions. Let  $\mathcal{Z} = \mathbb{R}^d$  denote the space in which the control process takes values.

**Definition 2.5.** For the three possible Regimes  $i = 1, 2, 3$  defined in (1.2) and for  $x \in \mathbb{R}^d, y \in \mathcal{Y}$  and  $z \in \mathcal{Z}$ , let

$$\begin{aligned} \mathcal{L}_x^1 &= b(x, y) \cdot \nabla_y + \frac{1}{2} \sigma(x, y) \sigma(x, y)^T : \nabla_y \nabla_y \\ \mathcal{L}_{z,x}^2 &= [\gamma b(x, y) + c(x, y) + \sigma(x, y)z] \cdot \nabla_y + \gamma \frac{1}{2} \sigma(x, y) \sigma(x, y)^T : \nabla_y \nabla_y \\ \mathcal{L}_{z,x}^3 &= [c(x, y) + \sigma(x, y)z] \cdot \nabla_y. \end{aligned}$$

For  $i = 1, 2$  we let  $\mathcal{D}(\mathcal{L}_{z,x}^i) = \mathcal{C}^2(\mathcal{Y})$  and for  $i = 3$ ,  $\mathcal{D}(\mathcal{L}_{z,x}^3) = \mathcal{C}^1(\mathcal{Y})$ .

We also define for Regime  $i$  a function  $\lambda_i(x, y, z)$ ,  $i = 1, 2, 3$ , as follows.

**Definition 2.6.** For the three possible Regimes  $i = 1, 2, 3$  defined in (1.2) and for  $x \in \mathbb{R}^d, y \in \mathcal{Y}$  and  $z \in \mathcal{Z}$ , define  $\lambda_i(x, y, z) : \mathbb{R}^d \times \mathcal{Y} \times \mathcal{Z} \rightarrow \mathbb{R}^d$  by

$$\begin{aligned}\lambda_1(x, y, z) &= \left( I + \frac{\partial \chi}{\partial y}(x, y) \right) (c(x, y) + \sigma(x, y)z) \\ \lambda_2(x, y, z) &= \gamma b(x, y) + c(x, y) + \sigma(x, y)z \\ \lambda_3(x, y, z) &= c(x, y) + \sigma(x, y)z,\end{aligned}$$

where  $\chi = (\chi_1, \dots, \chi_d)$  is defined by (2.1) and  $I$  is the identity matrix.

For a Polish space  $\mathcal{S}$ , let  $\mathcal{P}(\mathcal{S})$  be the space of probability measures on  $\mathcal{S}$ . Let  $\Delta = \Delta(\epsilon) \downarrow 0$  as  $\epsilon \downarrow 0$ . The role of  $\Delta(\epsilon)$  is to exploit a time-scale separation. Let  $A, B, \Gamma$  be Borel sets of  $\mathcal{Z}, \mathcal{Y}, [0, 1]$  respectively. Let  $u^\epsilon \in \mathcal{A}$  and let  $\bar{X}_s^\epsilon$  solve (2.2) with  $u^\epsilon$  in place of  $u$ . We associate with  $\bar{X}^\epsilon$  and  $u^\epsilon$  a family of occupation measures  $P^{\epsilon, \Delta}$  defined by

$$(2.3) \quad P^{\epsilon, \Delta}(A \times B \times \Gamma) = \int_\Gamma \left[ \frac{1}{\Delta} \int_t^{t+\Delta} 1_A(u_s^\epsilon) 1_B \left( \frac{\bar{X}_s^\epsilon}{\delta} \mod 1 \right) ds \right] dt,$$

with the convention that if  $s > 1$  then  $u_s^\epsilon = 0$ .

The first result, Theorem 2.8, deals with the limiting behavior of the controlled process (2.2) under each of the three regimes, and uses the notion of a viable pair.

**Definition 2.7.** A pair  $(\psi, P) \in \mathcal{C}([0, 1]; \mathbb{R}^d) \times \mathcal{P}(\mathcal{Z} \times \mathcal{Y} \times [0, 1])$  will be called viable with respect to  $(\lambda, \mathcal{L})$ , or simply viable if there is no confusion, if the following are satisfied. The function  $\psi_t$  is absolutely continuous,  $P$  is square integrable in the sense that  $\int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P(dz dy ds) < \infty$ , and the following hold for all  $t \in [0, 1]$ :

$$(2.4) \quad \psi_t = x_0 + \int_{\mathcal{Z} \times \mathcal{Y} \times [0, t]} \lambda(\psi_s, y, z) P(dz dy ds),$$

for every  $f \in \mathcal{D}(\mathcal{L})$

$$(2.5) \quad \int_0^t \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z, \psi_s} f(y) P(dz dy ds) = 0,$$

and

$$(2.6) \quad P(\mathcal{Z} \times \mathcal{Y} \times [0, t]) = t.$$

We write  $(\psi, P) \in \mathcal{V}_{(\lambda, \mathcal{L})}$  or simply  $(\psi, P) \in \mathcal{V}$  if there is no confusion.

Equation (2.6) implies that the last marginal of  $P$  is Lebesgue measure, and hence  $P$  can be decomposed in the form  $P(dz dy dt) = P_t(dz dy) dt$ . Equations (2.6) and (2.5) then imply that, for a choice of the kernel  $P_t(dz dy)$ ,  $P_t(\mathcal{Z} \times \mathcal{Y}) = 1$  and

$$\int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z, \psi_t} f(y) P_t(dz dy) = 0,$$

and by (2.4) for a.e.  $t \in [0, 1]$

$$\dot{\psi}_t = \int_{\mathcal{Z} \times \mathcal{Y}} \lambda(\psi_t, y, z) P_t(dz dy).$$

Note that a viable pair depends on the initial condition  $\psi_0 = x_0$  as well. Since this is will be deterministic and fixed throughout the paper, we frequently omit writing this dependence explicitly.

**Theorem 2.8.** Given  $x_0 \in \mathbb{R}^d$ , consider any family  $\{u^\epsilon, \epsilon > 0\}$  of controls in  $\mathcal{A}$  satisfying

$$\sup_{\epsilon > 0} \mathbb{E} \int_0^1 \|u_t^\epsilon\|^2 dt < \infty$$

and assume Condition 2.2. In addition, in Regime 1 assume Condition 2.3. Then the family  $\{(\bar{X}^\epsilon, P^{\epsilon, \Delta}), \epsilon > 0\}$  is tight. Hence given Regime  $i$ ,  $i = 1, 2, 3$ , and given any subsequence of  $\{(\bar{X}^\epsilon, P^{\epsilon, \Delta}), \epsilon > 0\}$ , there exists a subsubsequence that converges in distribution with limit  $(\bar{X}^i, P^i)$ . With probability 1, the accumulation point  $(\bar{X}^i, P^i)$  is a viable pair with respect to  $(\lambda_i, \mathcal{L}^i)$  according to Definition 2.7, i.e.,  $(\bar{X}^i, P^i) \in \mathcal{V}_{(\lambda_i, \mathcal{L}^i)}$ .

A proof is given in Section 3. The following theorem is the main result of this paper. It asserts that a large deviation principle holds, and gives a unifying expression for the rate function for all three regimes.

**Theorem 2.9.** Let  $\{X^\epsilon, \epsilon > 0\}$  be the unique strong solution to (1.1). Assume Condition 2.2 and that we are considering Regime  $i$ , where  $i = 1, 2, 3$ . In Regime 1 assume Condition 2.3 and in Regime 3 assume either that we are in dimension  $d = 1$ , or that  $c(x, y) = c(y)$  and  $\sigma(x, y) = \sigma(y)$  for the general multidimensional case. Define

$$(2.7) \quad S^i(\phi) = \inf_{(\phi, P) \in \mathcal{V}_{(\lambda_i, \mathcal{L}^i)}} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P(dz dy dt) \right],$$

with the convention that the infimum over the empty set is  $\infty$ . Then for every bounded and continuous function  $h$  mapping  $\mathcal{C}([0, 1]; \mathbb{R}^d)$  into  $\mathbb{R}$

$$\lim_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] = \inf_{\phi \in \mathcal{C}([0, 1]; \mathbb{R}^d)} [S^i(\phi) + h(\phi)].$$

Moreover, for each  $s < \infty$ , the set

$$\Phi_s^i = \{\phi \in \mathcal{C}([0, 1]; \mathbb{R}^d) : S^i(\phi) \leq s\}$$

is a compact subset of  $\mathcal{C}([0, 1]; \mathbb{R}^d)$ . In other words,  $\{X^\epsilon, \epsilon > 0\}$  satisfies the Laplace principle with rate function  $S^i$ .

The proof of this theorem is given in the subsequent sections. In Section 5 we prove that the formulation given in (2.7) for the rate function takes an explicit form in Regime 1 which agrees with the formula provided in [20]. We also construct a nearly optimal control that achieves the LDP lower bound (or equivalently the Laplace principle upper bound) at the prelimit level, see Theorem 5.3. In Sections 6 and 7, similar constructions are provided for Regimes 2 and 3, respectively.

**Remark 2.10.** In the case of Regime 3 we prove the Laplace principle lower bound for the general  $d$ -dimensional  $(x, y)$ -dependent case. However, for reasons that will be explained in Section 7, we can prove the Laplace principle upper bound for the general  $(x, y)$ -dependent case in dimension  $d = 1$  and under the assumption that  $c$  and  $\sigma$  are independent of  $x$  for the general multidimensional case. We conjecture that the full Laplace principle holds without this restriction, and note also that the rate function for Regime 3 is a limiting case of that of Regime 2 obtained by setting  $\gamma = 0$ .

**Remark 2.11.** The regularity assumptions imposed in Condition 2.2 can be relaxed. Due to Condition 2.2, the solution to the cell problem (2.1) is twice differentiable, which allows us to apply Itô's formula. Consider the case  $b = b(y)$ ,  $\sigma = \sigma(y)$  and assume that they are Lipschitz continuous. Then, standard elliptic regularity theory (e.g., [23]) shows that the solution  $\chi$  to equation (2.1) is in  $H^2(\mathcal{Y}) = W^{2,2}(\mathcal{Y})$ . By Sobolev's embedding lemma it is also in  $C^1(\mathcal{Y})$ . Then, using a standard approximation argument, one can still prove Theorems 2.8 and 2.9 for Regime 1.



We conclude this section with a remark on possible extensions of Theorem 2.9 to the case  $\mathcal{Y} = \mathbb{R}^d$ .

**Remark 2.12.** *In the case of Regime 1 and under some additional assumptions, one can extend the results to  $\mathcal{Y} = \mathbb{R}^d$ . In particular, one needs to impose structural assumptions on the coefficients  $b$  and  $\sigma$  such that an invariant measure corresponding to the operator  $\mathcal{L}_x^1$  exists. Also, note that for  $\mathcal{Y} = \mathbb{R}^d$  there are no boundary conditions associated with the cell problem (2.1). One looks for solutions that grow at most polynomially in  $y$ , as  $\|y\| \rightarrow \infty$ . For more details and specific statements on homogenization for fast oscillating diffusion processes on the whole space, see [36]. Using these results and techniques similar to the ones developed in the current paper, one can prove results that are analogous to Theorem 2.8 and Theorem 2.9 for Regime 1 and  $\mathcal{Y} = \mathbb{R}^d$ .*

*The situation is a bit more complicated for Regimes 2 and 3. One of the main reasons is that the operators  $\mathcal{L}_{z,x}^2$  and  $\mathcal{L}_{z,x}^3$  involve the control variable as well. However, using results on the structure of solutions to ergodic type Bellman equations in  $\mathbb{R}^d$  analogous to [26, 10] and techniques similar to the ones developed in the current paper, it seems possible that one can prove a result that is analogous to Theorem 2.9 for Regime 2 and  $\mathcal{Y} = \mathbb{R}^d$ . Ergodic type Bellman equations arise naturally in the study of the local rate function in Subsection 6.1. Assuming special structure on the dynamics, the authors in [31] and [41] have looked at similar problems corresponding to Regime 2 when  $\mathcal{Y} = \mathbb{R}^d$ , using other methods. Among other assumptions, the author in [31] assumes that the fast variable enters the equations of motion in an affine fashion, whereas the author in [41] assumes that the diffusion coefficient of the fast motion is independent of the slow motion. However, the arguments used in [31, 41] do not seem to directly extend to the full nonlinear case.*

### 3. LIMITING BEHAVIOR OF THE CONTROLLED PROCESS

In this section we prove Theorem 2.8. In particular, in Subsection 3.1 we prove tightness of the pair  $(\bar{X}^\epsilon, P^{\epsilon,\Delta})$  and in Subsection 3.2 we prove that any accumulation point  $(\bar{X}^i, P^i)$  of  $(\bar{X}^\epsilon, P^{\epsilon,\Delta})$  is a viable pair according to Definition 2.7 for Regimes  $i = 1, 2, 3$ . Note that the approach is the same for all three regimes. Therefore, we present the proof in detail for Regime 1 and for Regimes 2 and 3 only outline the differences.

**3.1. Tightness.** In this section we prove that the pair  $(\bar{X}^\epsilon, P^{\epsilon,\Delta})$  is tight. The proof is independent of the regime under consideration.

**Proposition 3.1.** *Consider any family  $\{u^\epsilon, \epsilon > 0\}$  of controls in  $\mathcal{A}$  satisfying*

$$(3.1) \quad \sup_{\epsilon > 0} \mathbb{E} \int_0^1 \|u_t^\epsilon\|^2 dt < \infty$$

*and assume Condition 2.2. In addition, in Regime 1 assume Condition 2.3. Then the following hold.*

- (i) *The family  $\{(\bar{X}^\epsilon, P^{\epsilon,\Delta}), \epsilon > 0\}$  is tight.*
- (ii) *The family  $\{P^{\epsilon,\Delta}, \epsilon > 0\}$  is uniformly integrable in the sense that*

$$\lim_{M \rightarrow \infty} \sup_{\epsilon > 0} \mathbb{E}_{x_0} \left[ \int_{\{z \in \mathcal{Z} : \|z\| > M\} \times \mathcal{Y} \times [0,1]} \|z\| P^{\epsilon,\Delta}(dz dy dt) \right] = 0.$$

*Proof.* (i). Tightness of the family  $\{\bar{X}^\epsilon\}$  is standard if we take into account the assumptions on the coefficients and the fact that the sequence of controls  $\{u^\epsilon, \epsilon > 0\}$  in  $\mathcal{A}$  satisfy (3.1). Some care is needed only for Regime 1, because of the presence of the unbounded drift term. Recall that  $\chi = (\chi_1, \dots, \chi_d)$  is one periodic in every direction in  $y$  and satisfies

$$\mathcal{L}_x^1 \chi_\ell(x, y) = -b_\ell(x, y), \quad \int_{\mathcal{Y}} \chi_\ell(x, y) \mu(dy|x) = 0, \quad \ell = 1, \dots, d.$$

Applying Itô's formula to  $\chi(x, x/\delta) = (\chi_1(x, x/\delta), \dots, \chi_d(x, x/\delta))$  with  $x = \bar{X}_t^\epsilon$ , we get

$$\begin{aligned}
(3.2) \quad \bar{X}_t^\epsilon &= x_0 + \int_0^t \left( I + \frac{\partial \chi}{\partial y} \right) \left( \bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta} \right) \left[ c \left( \bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta} \right) + \sigma \left( \bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta} \right) u_s^\epsilon \right] ds \\
&+ \int_0^t \left[ \epsilon \frac{\partial \chi}{\partial x} b + \delta \frac{\partial \chi}{\partial x} [c + \sigma u_s^\epsilon] + \epsilon \delta \frac{1}{2} \sigma \sigma^T : \frac{\partial^2 \chi}{\partial x^2} + \epsilon \frac{1}{2} \sigma \sigma^T : \frac{\partial^2 \chi}{\partial x \partial y} \right] \left( \bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta} \right) ds \\
&+ \sqrt{\epsilon} \int_0^t \left[ \left( I + \frac{\partial \chi}{\partial y} \right) \sigma + \delta \frac{\partial \chi}{\partial x} \sigma \right] \left( \bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta} \right) dW_s - \delta \left[ \chi \left( \bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta} \right) - \chi \left( x_0, \frac{x_0}{\delta} \right) \right].
\end{aligned}$$

From this representation, the boundedness of the coefficients and the second derivatives of  $\chi$  and assumption (3.1), it follows that for every  $\eta > 0$

$$\lim_{\rho \downarrow 0} \limsup_{\epsilon \downarrow 0} \mathbb{P}_{x_0} \left[ \sup_{|t_1 - t_2| < \rho, 0 \leq t_1 < t_2 \leq 1} |\bar{X}_{t_1}^\epsilon - \bar{X}_{t_2}^\epsilon| \geq \eta \right] = 0.$$

This implies the tightness of  $\{\bar{X}^\epsilon\}$ .

It remains to prove tightness of the occupation measures  $\{P^{\epsilon, \Delta}, \epsilon > 0\}$ . We claim that the function

$$g(r) = \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 r(dz dy dt), \quad r \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y} \times [0, 1])$$

is a tightness function, i.e., it is bounded from below and its level sets  $R_k = \{r \in \mathcal{P}(\mathbb{R}^d \times \mathcal{Y} \times [0, 1]) : g(r) \leq k\}$  are relatively compact for each  $k < \infty$ . To prove the relative compactness, observe that Chebyshev's inequality implies

$$\sup_{r \in R_k} r(\{(z, y) \in \mathcal{Z} \times \mathcal{Y} : \|z\| > M\} \times [0, 1]) \leq \sup_{r \in R_k} \frac{g(r)}{M^2} \leq \frac{k}{M^2}.$$

Hence,  $R_k$  is tight and thus relatively compact as a subset of  $\mathcal{P}$ .

Since  $g$  is a tightness function, by Theorem A.3.17 of [18] tightness of  $\{P^{\epsilon, \Delta}, \epsilon > 0\}$  will follow if we prove that

$$\sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0} [g(P^{\epsilon, \Delta})] < \infty.$$

However, by (3.1)

$$\begin{aligned}
\sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0} [g(P^{\epsilon, \Delta})] &= \sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0} \left[ \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P^{\epsilon, \Delta}(dz dy dt) \right] \\
&= \sup_{\epsilon \in (0, 1]} \mathbb{E}_{x_0} \int_0^1 \frac{1}{\Delta} \int_t^{t+\Delta} \|u_s^\epsilon\|^2 ds dt \\
&< \infty.
\end{aligned}$$

(ii). This follows from the last display and

$$\mathbb{E}_{x_0} \left[ \int_{\{z \in \mathcal{Z} : \|z\| > M\} \times \mathcal{Y} \times [0, 1]} \|z\| P^{\epsilon, \Delta}(dz dy dt) \right] \leq \frac{1}{M} \mathbb{E}_{x_0} \left[ \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P^{\epsilon, \Delta}(dz dy dt) \right].$$

This concludes the proof of the proposition.  $\square$



**3.2. Weak convergence analysis.** Before beginning the weak convergence analysis we make an observation that is useful in the proofs for all three regimes. Let

$$(3.3) \quad g(\epsilon) = \begin{cases} \frac{\delta^2}{\epsilon} & \text{Regime 1,} \\ \epsilon & \text{Regime 2,} \\ \delta & \text{Regime 3,} \end{cases}$$

where we recall that  $\delta = \delta(\epsilon) \downarrow 0$  as  $\epsilon \downarrow 0$ . Then the particular relation between  $\delta$  and  $\epsilon$  in each regime as given by (1.2) implies that  $g(\epsilon) \downarrow 0$  as  $\epsilon \downarrow 0$ . The process  $\bar{Y}_t^\epsilon = \bar{X}_t^\epsilon / \delta$  satisfies

$$(3.4) \quad \bar{Y}_t^\epsilon = \frac{x_0}{\delta} + \int_0^t \left[ \frac{\epsilon}{\delta^2} b(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon) + \frac{1}{\delta} c(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon) + \frac{1}{\delta} \sigma(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon) u_s^\epsilon \right] ds + \frac{\sqrt{\epsilon}}{\delta} \int_0^t \sigma(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon) dW_s.$$

Recall the operators  $\mathcal{L}_{z,x}^i$  for  $i = 1, 2, 3$  as given in Definition 2.5. Suppose that instead of (3.4) we consider the analogous equation with the slow motion and control “frozen,” i.e., with  $\bar{X}_s^\epsilon$  replaced by  $x$  and  $u_s^\epsilon$  replaced by  $z$ , and define  $\mathcal{A}_{z,x}^\epsilon$  by

$$(3.5) \quad \mathcal{A}_{z,x}^\epsilon f(y) = \left[ \frac{\epsilon}{\delta^2} b(x, y) + \frac{1}{\delta} [c(x, y) + \sigma(x, y)z] \right] \cdot \nabla_y f(y) + \frac{\epsilon}{\delta^2} \frac{1}{2} \sigma \sigma^T(x, y) : \nabla_y \nabla_y f(y)$$

for suitable functions  $f$ . Then it is easy to check that  $g(\epsilon) \mathcal{A}_{z,x}^\epsilon$  converges to  $\mathcal{L}_{z,x}^i$  under Regime  $i = 1, 3$  and to  $\gamma \mathcal{L}_{z,x}^2$  under Regime  $i = 2$ , as  $\epsilon \downarrow 0$ .

**3.2.1. Limiting behavior of the CSDE in Regime 1.** In this section we prove Theorem 2.8 for  $i = 1$ . For notational convenience we drop the subscript or superscript 1 from  $\lambda_1, \bar{X}^1$  and  $P^1$ .

**Lemma 3.2.** *Let  $T > 0$  and  $\tau > 0$  be positive numbers such that  $T + \tau \leq 1$ . Consider a continuous function  $g : \mathbb{R}^d \times \mathcal{Y} \times \mathcal{Z} \rightarrow \mathbb{R}$  that is bounded in the first and the second argument and affine in the third argument. Assume that  $(\bar{X}^\epsilon, P^{\epsilon, \Delta}) \rightarrow (\bar{X}, P)$  in distribution for some subsequence of  $\epsilon \downarrow 0$ , and that Conditions 2.2 and 2.3 and (3.1) hold. Then the following limits are valid in distribution along this subsequence:*

$$(3.6) \quad \int_{\mathcal{Z} \times \mathcal{Y} \times [T, T+\tau]} g(\bar{X}_t^\epsilon, y, z) P^{\epsilon, \Delta}(dz dy dt) \rightarrow \int_{\mathcal{Z} \times \mathcal{Y} \times [T, T+\tau]} g(\bar{X}_t, y, z) P(dz dy dt)$$

and

$$(3.7) \quad \int_T^{T+\tau} g\left(\bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta}, u_t^\epsilon\right) dt - \int_{\mathcal{Z} \times \mathcal{Y} \times [T, T+\tau]} g(\bar{X}_t^\epsilon, y, z) P^{\epsilon, \Delta}(dz dy dt) \rightarrow 0.$$

*Proof.* First note that (3.6) holds due to the weak convergence, the fact that the last marginal of  $P^{\epsilon, \Delta}(dz dy dt)$  is always Lebesgue measure and part (ii) of Proposition 3.1 (see [15], page 137 for more details).

Next we show that (3.7) holds. This follows from the following three observations.

- (i) Change of the order of integration implies that if  $\tilde{h}(s) : [0, \infty) \rightarrow \mathbb{R}$  is integrable on each bounded interval then

$$(3.8) \quad \left| \int_0^T \frac{1}{\Delta} \int_t^{t+\Delta} \tilde{h}(s) ds dt - \int_0^T \tilde{h}(s) ds \right| \leq \int_0^\Delta |\tilde{h}(s)| ds + \int_T^{T+\Delta} |\tilde{h}(s)| ds.$$

- (ii) The definition of the occupation measure  $P^{\epsilon, \Delta}$  gives

$$(3.9) \quad \int_{\mathcal{Z} \times \mathcal{Y} \times [T, T+\tau]} g(\bar{X}_t^\epsilon, y, z) P^{\epsilon, \Delta}(dz dy dt) = \int_T^{T+\tau} \frac{1}{\Delta} \int_t^{t+\Delta} g\left(\bar{X}_t^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta}, u_s^\epsilon\right) ds.$$

(iii) The tightness of  $\{\bar{X}^\epsilon\}$  implies that for every  $\eta > 0$

$$\sup_{0 \leq t \leq T} \mathbb{P}_{x_0} \left[ \sup_{0 \leq t \leq t+\Delta \leq T} |\bar{X}_{t+\Delta}^\epsilon - \bar{X}_t^\epsilon| > \eta \right] \rightarrow 0 \text{ as } \epsilon \downarrow 0.$$

The last display, together with the continuity of  $g$  in the first variable, the fact that the second variable takes values in a compact space, and part (ii) of Proposition 3.1, imply that

$$\int_T^{T+\tau} \frac{1}{\Delta} \int_t^{t+\Delta} g\left(\bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta}, u_s^\epsilon\right) ds dt - \int_T^{T+\tau} \frac{1}{\Delta} \int_t^{t+\Delta} g\left(\bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta}, u_s^\epsilon\right) ds dt \rightarrow 0 \text{ in probability.}$$

Then (3.9), the last display, (3.8) and (3.1) show that (3.7) holds.  $\square$

*Proof of Theorem 2.8 for  $i = 1$ .* The tightness proven in Proposition 3.1 implies that for any subsequence of  $\epsilon > 0$  there exists a convergent subsubsequence and  $(\bar{X}, P)$  such that

$$(\bar{X}^\epsilon, P^{\epsilon, \Delta}) \rightarrow (\bar{X}, P) \text{ in distribution.}$$

We invoke the Skorokhod representation theorem (Theorem 1.8 in [19]) which allows us to assume that the aforementioned convergence holds with probability 1. The Skorokhod representation theorem involves the introduction of another probability space, but this distinction is ignored in the notation. Note that by Fatou's Lemma

$$(3.10) \quad \mathbb{E}_{x_0} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P(dz dy dt) < \infty$$

and so  $\int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P(dz dy dt) < \infty$  w.p.1. Thus it remains to show that  $(\bar{X}, P)$  satisfy (2.4), (2.5) and (2.6).

Our tool for proving (2.4) will be the characterization of solutions to SDE's via the martingale problem [19]. Let  $f, \phi_j$  be smooth, real valued functions with compact support. For a measure  $r \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y} \times [0,1])$  and  $t \in [0,1]$ , define

$$(r, \phi_j)_t = \int_{\mathcal{Z} \times \mathcal{Y} \times [0,t]} \phi_j(z, y, s) r(dz dy ds).$$

Let  $T, t_i, \tau \geq 0, i \leq q$  be given such that  $t_i \leq T \leq T + \tau \leq 1$  and let  $\zeta$  be a real valued, bounded and continuous function with compact support on  $(\mathbb{R}^d)^q \times \mathbb{R}^{pq}$ . We recall that

$$\lambda(x, y, z) = \left( I + \frac{\partial \chi}{\partial y}(x, y) \right) (c(x, y) + \sigma(x, y)z).$$

In order to prove (2.4), it is sufficient to prove for any fixed such collection  $p, q, T, t_i, \tau, \phi_j, \zeta, f$  that, as  $\epsilon \downarrow 0$ ,

$$(3.11) \quad \mathbb{E}_{x_0} \left[ \zeta(\bar{X}_{t_i}^\epsilon, (P^{\epsilon, \Delta}, \phi_j)_{t_i}, i \leq q, j \leq p) \left[ f(\bar{X}_{T+\tau}^\epsilon) - f(\bar{X}_T^\epsilon) - \int_T^{T+\tau} \bar{\mathcal{A}}_t^{\epsilon, \Delta} f(\bar{X}_t^\epsilon) dt \right] \right] \rightarrow 0$$

and

$$(3.12) \quad \int_T^{T+\tau} \bar{\mathcal{A}}_s^{\epsilon, \Delta} f(\bar{X}_s^\epsilon) ds - \int_{\mathcal{Z} \times \mathcal{Y} \times [T, T+\tau]} \lambda(\bar{X}_s, y, z) \nabla f(\bar{X}_s) P(dz dy ds) \rightarrow 0$$

in probability. Here  $\bar{\mathcal{A}}_s^{\epsilon, \Delta}$  is defined by

$$(3.13) \quad \bar{\mathcal{A}}_t^{\epsilon, \Delta} f(x) = \int_{\mathcal{Z} \times \mathcal{Y}} \lambda(x, y, z) \nabla f(x) P_t^{\epsilon, \Delta}(dz dy)$$

and

$$P_t^{\epsilon, \Delta}(dz dy) = \frac{1}{\Delta} \int_t^{t+\Delta} 1_{dz}(u_s^\epsilon) 1_{dy} \left( \frac{\bar{X}_s^\epsilon}{\delta} \bmod 1 \right) ds.$$

Since they show that  $(\bar{X}, P)$  solves the appropriate martingale problem, relations (3.11) and (3.12) imply (2.4). So, let us prove now that (3.11) and (3.12) hold.

First, for every real valued, continuous function  $\phi$  with compact support and  $t \in [0, 1]$

$$(P^{\epsilon, \Delta}, \phi)_t \rightarrow (P, \phi)_t \text{ w.p.1.}$$

This follows from the topology used and the fact that the last marginal of  $P$  is Lebesgue measure w.p.1. Second, we recall the solution  $\chi(x, y)$  to the cell problem (2.1) and consider the function  $\psi_\ell(x, y) = \chi_\ell(x, y)f_{x_\ell}(x)$  for  $\ell = 1, \dots, d$ . Then  $\psi_\ell(x, y)$  is one periodic in every direction in  $y$  and satisfies

$$(3.14) \quad \mathcal{L}_x^1 \psi_\ell(x, y) = -b_\ell(x, y)f_{x_\ell}(x), \quad \int_{\mathcal{Y}} \psi_\ell(x, y)\mu(dy|x) = 0.$$

Let  $\psi = \{\psi_1, \dots, \psi_d\}$ . We apply Itô's formula to  $\psi(\bar{X}_s^\epsilon, \bar{X}_s^\epsilon/\delta)$ . Relation (3.14) and the boundedness of  $\chi(x, y)$  and its derivatives (see (3.2)) imply that in order to show (3.11), it is sufficient to show that

$$(3.15) \quad \int_T^{T+\tau} \left[ \bar{\mathcal{A}}_s^{\epsilon, \Delta} f(\bar{X}_s^\epsilon) - \lambda \left( \bar{X}_s^\epsilon, \frac{\bar{X}_s^\epsilon}{\delta}, u_s^\epsilon \right) \nabla f(\bar{X}_s^\epsilon) \right] ds \rightarrow 0, \text{ as } \epsilon \downarrow 0.$$

in probability (a number of other terms converge to zero and we do not write them explicitly for notational convenience). However, we can apply Lemma 3.2 to

$$g(x, y, z) = \lambda(x, y, z) \cdot \nabla f(x) = \left( \left( I + \frac{\partial \chi}{\partial y}(x, y) \right) c(x, y) + \left( I + \frac{\partial \chi}{\partial y}(x, y) \right) \sigma(x, y)z \right) \cdot \nabla f(x),$$

in which case (3.12) follows from (3.6), and also (3.15) (and hence (3.11)) follows from (3.7). The completes the proof of (2.4).

Next we prove that (2.5) holds. For this purpose define  $\bar{Y}^\epsilon = \bar{X}^\epsilon/\delta$ . Let  $f_\ell : \mathcal{Y} \mapsto \mathbb{R}$ ,  $\ell \in \mathbb{N}$  be smooth and dense in  $\mathcal{C}(\mathcal{Y})$ . Observe that the quantity

$$(3.16) \quad M_t^\epsilon = f_\ell(\bar{Y}_t^\epsilon) - f_\ell(x_0/\delta) - \int_0^t \mathcal{A}_{u_s^\epsilon, \bar{X}_s^\epsilon}^\epsilon f_\ell(\bar{Y}_s^\epsilon) ds = \frac{\sqrt{\epsilon}}{\delta} \int_0^t \nabla_y f_\ell(\bar{Y}_s^\epsilon) \cdot \sigma(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon) dW_s,$$

where  $\mathcal{A}_{z,x}^\epsilon$  is defined in (3.5), is an  $\mathfrak{F}_t$ -martingale. Moreover, for any  $T > 0$ , we have from (3.8) that

$$\begin{aligned} \int_0^T \mathcal{A}_{u_t^\epsilon, \bar{X}_t^\epsilon}^\epsilon f_\ell(\bar{Y}_t^\epsilon) dt + e_T^\epsilon &= \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \mathcal{A}_{u_s^\epsilon, \bar{X}_s^\epsilon}^\epsilon f_\ell(\bar{Y}_s^\epsilon) ds \right] dt \\ &= \frac{\epsilon}{\delta^2} \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \mathcal{L}_{\bar{X}_s^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) ds \right] dt \\ &\quad + \frac{1}{\delta} \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} [c(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon) + \sigma(\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon)u_s^\epsilon] \cdot \nabla_y f_\ell(\bar{Y}_s^\epsilon) ds \right] dt, \end{aligned}$$

where

$$|e_T^\epsilon| \leq \int_0^\Delta |\mathcal{A}_{u_s^\epsilon, \bar{X}_s^\epsilon}^\epsilon f_\ell(\bar{Y}_s^\epsilon)| ds + \int_T^{T+\Delta} |\mathcal{A}_{u_s^\epsilon, \bar{X}_s^\epsilon}^\epsilon f_\ell(\bar{Y}_s^\epsilon)| ds.$$

Recall now the definition  $g(\epsilon) = \delta^2/\epsilon \rightarrow 0$  from (3.3) and define the operator

$$\mathcal{G}_{x,y,z} f_\ell(y) = [c(x, y) + \sigma(x, y)z] \cdot \nabla_y f_\ell(y).$$

Let  $D \subset [0, 1]$  be countable and dense, and consider any  $T \in D$  and  $\ell \in \mathbb{N}$ . By (3.16)

$$\begin{aligned}
& g(\epsilon)M_T^\epsilon - g(\epsilon) [f_\ell(\bar{Y}_T^\epsilon) - f_\ell(\bar{Y}_0^\epsilon)] + g(\epsilon)e_T^\epsilon \\
&= \frac{1}{\delta} \int_0^T \frac{g(\epsilon)}{\Delta} \left[ \int_t^{t+\Delta} \mathcal{G}_{\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon, u_s^\epsilon} f_\ell(\bar{Y}_s^\epsilon) ds \right] dt + \frac{\epsilon}{\delta^2} \int_0^T \frac{g(\epsilon)}{\Delta} \left[ \int_t^{t+\Delta} \mathcal{L}_{\bar{X}_s^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) ds \right] dt \\
&= \frac{g(\epsilon)}{\delta} \left( \int_0^T \frac{1}{\Delta} \int_t^{t+\Delta} \left[ \mathcal{G}_{\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon, u_s^\epsilon} f_\ell(\bar{Y}_s^\epsilon) - \mathcal{G}_{\bar{X}_t^\epsilon, \bar{Y}_s^\epsilon, u_s^\epsilon} f_\ell(\bar{Y}_s^\epsilon) \right] ds dt \right) \\
&+ \frac{g(\epsilon)}{\delta} \left( \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \mathcal{G}_{\bar{X}_t^\epsilon, \bar{Y}_s^\epsilon, u_s^\epsilon} f_\ell(\bar{Y}_s^\epsilon) ds \right] dt \right) \\
&+ \frac{\epsilon g(\epsilon)}{\delta^2} \left( \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \left[ \mathcal{L}_{\bar{X}_s^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) - \mathcal{L}_{\bar{X}_t^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) \right] ds \right] dt \right) \\
&+ \frac{\epsilon g(\epsilon)}{\delta^2} \left( \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \left[ \mathcal{L}_{\bar{X}_t^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) \right] ds \right] dt \right) \\
&= \frac{\delta}{\epsilon} \left( \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \left[ \mathcal{G}_{\bar{X}_s^\epsilon, \bar{Y}_s^\epsilon, u_s^\epsilon} f_\ell(\bar{Y}_s^\epsilon) - \mathcal{G}_{\bar{X}_t^\epsilon, \bar{Y}_s^\epsilon, u_s^\epsilon} f_\ell(\bar{Y}_s^\epsilon) \right] ds \right] dt \right) \\
&+ \frac{\delta}{\epsilon} \left( \int_{\mathcal{Z} \times \mathcal{Y} \times [0, T]} \mathcal{G}_{\bar{X}_t^\epsilon, y, z} f_\ell(y) P^{\epsilon, \Delta}(dz dy dt) \right) \\
&+ \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \left[ \mathcal{L}_{\bar{X}_s^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) - \mathcal{L}_{\bar{X}_t^\epsilon}^1 f_\ell(\bar{Y}_s^\epsilon) \right] ds \right] dt \\
(3.17) \quad &+ \int_{\mathcal{Z} \times \mathcal{Y} \times [0, T]} \mathcal{L}_{\bar{X}_t^\epsilon}^1 f_\ell(y) P^{\epsilon, \Delta}(dz dy dt).
\end{aligned}$$

First consider the left hand side of (3.17). Since  $f_\ell$  is bounded  $g(\epsilon) [f_\ell(\bar{Y}_T^\epsilon) - f_\ell(\bar{Y}_0^\epsilon)]$  converges to zero uniformly. We claim that

$$(3.18) \quad g(\epsilon)M_T^\epsilon \rightarrow 0 \text{ in probability.}$$

Indeed, since  $\sigma$  is uniformly bounded  $\mathbb{E}_{x_0} [M_T^\epsilon]^2$  is bounded above by a constant times  $\epsilon/\delta^2 = 1/g(\epsilon)$ , and so (3.18) also follows from  $g(\epsilon) \downarrow 0$ . Finally, we claim that  $g(\epsilon)e_T^\epsilon \rightarrow 0$  in probability. Using Condition 2.2, for some constants  $C_1$  and  $C_2$

$$\begin{aligned}
g(\epsilon) \int_0^\Delta |\mathcal{A}_{u_s^\epsilon, \bar{X}_s^\epsilon}^\epsilon f_\ell(\bar{Y}_s^\epsilon)| ds &\leq g(\epsilon)C_1 \Delta \frac{\epsilon}{\delta^2} + g(\epsilon)C_2 \frac{1}{\delta} \int_0^\Delta (1 + \|u_s^\epsilon\|) ds \\
&\leq C_1 \Delta + C_2 \frac{\delta}{\epsilon} \left[ \frac{3}{2} \Delta + \frac{1}{2} \int_0^1 \|u_s^\epsilon\|^2 ds \right],
\end{aligned}$$

and hence the left hand side tends to zero in probability by (3.1) and since  $\Delta \downarrow 0, \delta/\epsilon \downarrow 0$ . The same estimate holds for the second term in  $g(\epsilon)e_T^\epsilon$ , and so the claim follows.

Next consider the right hand side of (3.17). The first and the third term in the right hand side of (3.17) converge to zero in probability by the tightness of  $\bar{X}^\epsilon$ , Condition 2.2, (3.1) and  $\delta/\epsilon \downarrow 0$ . The second term on the right hand side of (3.17) converges to zero in probability by the uniform integrability of  $P^{\epsilon, \Delta}$  and by the fact that  $\delta/\epsilon \downarrow 0$ . So, it remains to consider the fourth term. Passing to the limit as  $\epsilon \downarrow 0$ , the previous discussion implies that except on a set  $N_{\ell, T}$  of probability zero,

$$(3.19) \quad 0 = \int_0^T \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{\bar{X}_t}^1 f_\ell(y) P(dz dy dt).$$

Let  $N = \cup_{\ell \in \mathbb{N}} \cup_{T \in D} N_{\ell, T}$ . Then except on the set  $N$  of probability zero, continuity in  $T$  and denseness of  $\{f_\ell, \ell \in \mathbb{N}\}$  imply that (3.19) holds for all  $T \in [0, 1]$  and all  $f \in \mathcal{C}^2(\mathcal{Y})$ .

It remains to prove that  $P(\mathcal{Z} \times \mathcal{Y} \times [0, t]) = t$  for every  $t \in [0, 1]$ . Using the fact that the analogous property holds at the prelimit level,  $P(\mathcal{Z} \times \mathcal{Y} \times \{t\}) = 0$  and the continuity of  $t \rightarrow P(\mathcal{Z} \times \mathcal{Y} \times [0, t])$  to deal with null sets, this property also follows.  $\square$

**3.2.2. Limiting behavior of the CSDE in Regimes 2 and 3.** In this subsection we prove Theorem 2.8 for  $i = 2$ . The proof for  $i = 3$  is similar and thus it is omitted.

*Proof of Theorem 2.8 for  $i = 2$ .* The proof follows the same steps as the proof of Theorem 2.8 for  $i = 1$ , and hence only the differences are outlined. We have

$$\lambda(x, y, z) = \gamma b(x, y) + c(x, y) + \sigma(x, y)z$$

and the operator  $\bar{\mathcal{A}}_t^{\epsilon, \Delta}$  is defined as in (3.13), but with this particular function  $\lambda$ .

The proof of (2.4) can be carried out repeating the corresponding steps of the proof of Theorem 2.8 for  $i = 1$ . A difference is that one skips the step of applying Itô's formula to  $\psi_\ell$  that satisfies (3.14), since in this case we do not have an unbounded drift term.

It remains to discuss (2.5). Again, define  $\bar{Y}^\epsilon = \bar{X}^\epsilon / \delta$  and observe that for  $f_\ell : \mathcal{Y} \mapsto \mathbb{R}$ ,  $\ell \in \mathbb{N}$  smooth and dense in  $\mathcal{C}(\mathcal{Y})$ ,  $M_t^\epsilon$  defined by (3.16) is an  $\mathfrak{F}_t$ -martingale. For any  $T > 0$ , small  $\Delta > 0$  and recalling that in this case  $g(\epsilon) = \epsilon$ ,

$$\begin{aligned} & g(\epsilon)M_T^\epsilon - g(\epsilon) [f_\ell(\bar{Y}_T^\epsilon) - f_\ell(\bar{Y}_0^\epsilon)] + g(\epsilon)\epsilon T \\ &= \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \left( \epsilon \mathcal{A}_{u_s, \bar{X}_s}^\epsilon - \gamma \mathcal{L}_{u_s, \bar{X}_s}^2 \right) f_\ell(\bar{Y}_s^\epsilon) ds \right] dt + \gamma \int_0^T \frac{1}{\Delta} \left[ \int_t^{t+\Delta} \mathcal{L}_{u_s, \bar{X}_s}^2 f_\ell(\bar{Y}_s^\epsilon) ds \right] dt. \end{aligned}$$

Observing that the operator  $\epsilon \mathcal{A}_{z, x}^\epsilon$  converges to the operator  $\gamma \mathcal{L}_{z, x}^2$ , we can argue similarly to the corresponding part of the proof of Theorem 2.8 for  $i = 1$  and conclude that

$$\int_0^T \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z, \bar{X}_t}^2 f_\ell(y) P^2(dz dy dt) = 0 \text{ w.p.1.}$$

#### 4. LAPLACE PRINCIPLE LOWER BOUND AND COMPACTNESS OF LEVEL SETS

In this section we prove the Laplace principle lower bound for Theorem 2.9 and the compactness of the level sets of the action functional.

**4.1. Laplace principle lower bound.** For each  $\epsilon > 0$ , let  $X^\epsilon$  be the unique strong solution to (1.1). To prove the Laplace principle lower bound we must show that for all bounded, continuous functions  $h$  mapping  $\mathcal{C}([0, 1]; \mathbb{R}^d)$  into  $\mathbb{R}$

$$\liminf_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \geq \inf_{(\phi, P) \in \mathcal{V}} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P(dz dy dt) + h(\phi) \right].$$

Of course, it is sufficient to prove the lower limit (4.1) along any subsequence such that

$$-\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right]$$

converges. Such a subsequence exists since  $|\epsilon \ln \mathbb{E}_{x_0} [\exp \{-h(X^\epsilon)/\epsilon\}]| \leq \|h\|_\infty$ .

According to Theorem 2.4, there exists a family of controls  $\{u^\epsilon, \epsilon > 0\}$  in  $\mathcal{A}$  such that for every  $\epsilon > 0$

$$-\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \geq \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|u_t^\epsilon\|^2 dt + h(\bar{X}^\epsilon) \right] - \epsilon,$$

where the controlled process  $\bar{X}^\epsilon$  is defined in (2.2). Note that for each  $\epsilon > 0$   $\mathbb{E}_{x_0} \int_0^1 \|u_t^\epsilon\|^2 dt \leq 4 \|h\|_\infty + 2\epsilon$ , and hence if we use this family of controls and the associated controlled process  $\bar{X}^\epsilon$  to

construct occupation measures  $P^{\epsilon, \Delta}$  in (2.3), then by Proposition 3.1 the family  $\{\bar{X}^\epsilon, P^{\epsilon, \Delta}, \epsilon > 0\}$  is tight. Thus given any subsequence of  $\epsilon > 0$  there is a further subsubsequence for which

$$(\bar{X}^\epsilon, P^{\epsilon, \Delta}) \rightarrow (\bar{X}, P) \text{ in distribution}$$

with  $(\bar{X}, P) \in \mathcal{V}$ . By Fatou's lemma

$$\begin{aligned} \liminf_{\epsilon \downarrow 0} \left( -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \right) &\geq \liminf_{\epsilon \downarrow 0} \left( \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|u_t^\epsilon\|^2 dt + h(\bar{X}^\epsilon) \right] - \epsilon \right) \\ &\geq \liminf_{\epsilon \downarrow 0} \left( \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \frac{1}{\Delta} \int_t^{t+\Delta} \|u_s^\epsilon\|^2 ds dt + h(\bar{X}^\epsilon) \right] \right) \\ &= \liminf_{\epsilon \downarrow 0} \left( \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^{\epsilon, \Delta}(dz dy dt) + h(\bar{X}^\epsilon) \right] \right) \\ &\geq \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P(dz dy dt) + h(\bar{X}) \right] \\ &\geq \inf_{(\phi, P) \in \mathcal{V}} \left\{ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P(dz dy dt) + h(\phi) \right\}. \end{aligned}$$

This concludes the proof of the Laplace principle lower bound.

**4.2. Compactness of level sets.** Consider  $S^i(\phi)$  as defined by (2.7) and for notational convenience omit the superscript  $i$  since the proof is independent of the regime under consideration. We want to prove that for each  $s < \infty$ , the set

$$\Phi_s = \{\phi \in \mathcal{C}([0, 1]; \mathbb{R}^d) : S(\phi) \leq s\}$$

is a compact subset of  $\mathcal{C}([0, 1]; \mathbb{R}^d)$ . As usual with the weak convergence approach, the proof is analogous to that of the Laplace principle lower bound. In Lemma 4.1 we show precompactness of  $\Phi_s$  and in Lemma 4.3 that it is closed. Together they imply compactness of  $\Phi_s$ .

**Lemma 4.1.** *Fix  $K < \infty$  and consider any sequence  $\{(\phi^n, P^n), n > 0\}$  such that for every  $n > 0$   $(\phi^n, P^n)$  is viable and*

$$\int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^n(dz dy dt) < K.$$

*Then  $\{(\phi^n, P^n), n > 0\}$  is precompact.*

*Proof.* For any  $P$  such that  $(\phi, P) \in \mathcal{V}$

$$\begin{aligned} |\phi_{t_2} - \phi_{t_1}| &= \left| \int_{\mathcal{Z} \times \mathcal{Y} \times [t_1, t_2]} \lambda(\phi_s, y, z) P(dz dy ds) \right| \\ &\leq C_0 \left[ |t_2 - t_1| + \sqrt{(t_2 - t_1)} \sqrt{\int_{\mathcal{Z} \times \mathcal{Y} \times [t_1, t_2]} \|z\|^2 P(dz dy dt)} \right]. \end{aligned}$$

This implies the precompactness of  $\{\phi^n, n > 0\}$ . Precompactness of  $\{P^n, n > 0\}$  follows from the compactness of  $\mathcal{Y} \times [0, 1]$  and that  $\|z\|^2$  is a tightness function (similarly to Proposition 3.1, part (i)).  $\square$

Next, we prove that the limit of a viable pair is also viable.



**Lemma 4.2.** Fix  $K < \infty$  and consider any convergent sequence  $\{(\phi^n, P^n), n > 0\}$  such that for every  $n > 0$   $(\phi^n, P^n)$  is viable and

$$(4.1) \quad \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^n(dzdydt) < K.$$

Then  $(\phi, P)$  is a viable pair.

*Proof.* Since  $(\phi^n, P^n)$  is viable

$$(4.2) \quad \phi_t^n = x_0 + \int_{\mathcal{Z} \times \mathcal{Y} \times [0,t]} \lambda(\phi_s^n, y, z) P^n(dzdyds)$$

and

$$(4.3) \quad \int_0^t \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z, \phi_s^n} f(y) P^n(dzdyds) = 0$$

for every  $t \in [0, 1]$  and for every  $f \in \mathcal{C}^2(\mathcal{Y})$ . The function  $\lambda(x, y, z)$  and the operator  $\mathcal{L}_{z,x}$  are defined in Definitions 2.6 and 2.5 respectively.

By Fatou's Lemma  $P$  has a finite second moment in  $z$ . Moreover, observe that the function  $\lambda(x, y, z)$  and the operator  $\mathcal{L}_{z,x}$  are continuous in  $x$  and  $y$  and affine in  $z$ . Hence by assumption (4.1) and the convergence  $P^n \rightarrow P$  and  $\phi^n \rightarrow \phi$ ,  $(\phi, P)$  satisfy equation (4.2) with  $(\phi^n, P^n)$  replaced by  $(\phi, P)$ .

Next we show that (4.3) holds with  $(\phi^n, P^n)$  replaced by  $(\phi, P)$ . Since (4.1) holds and  $P(\mathcal{Z} \times \mathcal{Y} \times \{t\}) = 0$ , we can send  $n \rightarrow \infty$  in (4.3) and obtain

$$0 = \int_0^t \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z, \phi_s} f(y) P(dzdyds).$$

Finally, it follows from  $P^n(\mathcal{Z} \times \mathcal{Y} \times [0, t]) = t$  and  $P(\mathcal{Z} \times \mathcal{Y} \times \{t\}) = 0$  that  $P(\mathcal{Z} \times \mathcal{Y} \times [0, t]) = t$  for all  $t \in [0, 1]$ .  $\square$

**Lemma 4.3.** The functional  $S(\phi)$  is lower semicontinuous.

*Proof.* Let us consider a sequence  $\phi^n$  with limit  $\phi$ . We want to prove

$$\liminf_{n \rightarrow \infty} S(\phi^n) \geq S(\phi).$$

It suffices to consider the case when  $S(\phi^n)$  has a finite limit, i.e., there exists a  $M < \infty$  such that  $\liminf_{n \rightarrow \infty} S(\phi^n) \leq M$ .

We recall the definition

$$S(\phi^n) = \inf_{(\phi^n, P^n) \in \mathcal{V}_{(\lambda, \mathcal{L})}} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^n(dzdydt) \right].$$

Hence we can find measures  $\{P^n, n < \infty\}$  satisfying  $(\phi^n, P^n) \in \mathcal{V}_{(\lambda, \mathcal{L})}$  and

$$\sup_{n < \infty} \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^n(dzdyds) < M + 1,$$

and such that

$$S(\phi^n) \geq \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^n(dzdydt) - \frac{1}{n} \right].$$

It follows from Lemma 4.1 that we can consider a subsequence along which  $(\phi^n, P^n)$  converges to a limit  $(\phi, P)$ . By Lemma 4.2  $(\phi, P)$  is viable. Hence by Fatou's Lemma

$$\begin{aligned} \liminf_{n \rightarrow \infty} S(\phi^n) &\geq \liminf_{n \rightarrow \infty} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P^n(dzdydt) - \frac{1}{n} \right] \\ &\geq \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P(dzdydt) \\ &\geq \inf_{(\phi, P) \in \mathcal{V}(\lambda, \mathcal{L})} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0,1]} \|z\|^2 P(dzdydt) \right] \\ &= S(\phi), \end{aligned}$$

which concludes the proof of lower-semicontinuity of  $S(\cdot)$ .  $\square$

## 5. REGIME 1: LAPLACE PRINCIPLE UPPER BOUND AND ALTERNATIVE REPRESENTATION

In this section we prove the Laplace principle upper bound for Regime 1. We also prove in Theorem 5.3 that the formula for the rate function of Theorem 2.9 takes an explicit form which coincides with the form provided in [20]. For notational convenience we drop the superscript 1 from  $\bar{X}^1$  and  $P^1$ .

In each regime the same steps are taken. The rate function on path space obtained in the proof of the large deviation upper bound is defined in terms a viable pair through (2.7). What differs between regimes are the forms that  $\lambda_i$  and  $\mathcal{L}^i$  take. In each case, we consider for the limit variational problem in the Laplace principle a nearly optimal pair  $(\phi, P)$ . Using the notion of viability appropriate to the particular regime, we examine the constraints that link  $\phi$  and  $P$ . The last step is to construct, based on these constraints, a control for the prelimit representation that will lead to controls and controlled processes that will converge to the cost associated with  $P$  and  $\phi$ , respectively. This construction is subtle in all regimes, due to the multiscale aspect of the dynamics.

To begin the construction, first observe that one can write (2.7) in terms of a local rate function, i.e., in the form

$$S^1(\phi) = \int_0^1 L_1^r(\phi_s, \dot{\phi}_s) ds.$$

This follows from the definition of a viable pair by setting

$$(5.1) \quad L_1^r(x, \beta) = \inf_{P \in \mathcal{A}_{x, \beta}^{1, r}} \int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P(dzdy),$$

where

$$\mathcal{A}_{x, \beta}^{1, r} = \left\{ P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y}) : \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_x^1 f(y) P(dzdy) = 0 \text{ for all } f \in C^2(\mathcal{Y}) \right. \\ \left. \int_{\mathcal{Z} \times \mathcal{Y}} \|z\|^2 P(dzdy) < \infty \text{ and } \beta = \int_{\mathcal{Z} \times \mathcal{Y}} \lambda_1(x, y, z) P(dzdy) \right\}.$$

Note that any measure  $P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y})$  can be decomposed in the form

$$(5.2) \quad P(dzdy) = \eta(dz|y)\mu(dy),$$

where  $\mu$  is a probability measure on  $\mathcal{Y}$  and  $\eta$  is a stochastic kernel on  $\mathcal{Z}$  given  $\mathcal{Y}$ . We refer to this as a “relaxed” formulation because the control is characterized as a distribution on  $\mathcal{Z}$  (given  $x$  and

$y$ ) rather than as an element of  $\mathcal{Z}$ . Inserting (5.2) into (2.5) with  $\mathcal{L}_{z,x} = \mathcal{L}_x^1$  from Definition 2.5, we get that for every  $f \in \mathcal{C}^2(\mathcal{Y})$

$$(5.3) \quad \int_{\mathcal{Y}} \mathcal{L}_x^1 f(y) \mu(dy) = 0.$$

Here we have used the independence of  $\mathcal{L}_x^1$  on the control variable  $z$  to eliminate  $\eta$ . The nondegeneracy of the diffusion matrix  $\sigma\sigma^T$  and (5.3) guarantee that  $\mu(dy)$  is actually the unique invariant measure corresponding to the operator  $\mathcal{L}_x^1$  with periodic boundary conditions. Naturally,  $\mu(dy)$  implicitly depends on  $x$  and was identified in Condition 2.3 as  $\mu(dy|x)$ .

We note that because the cost is convex in  $z$  and  $\lambda_1$  is affine in  $z$ , the relaxed control formulation as given in (5.1) is equivalent to the following ordinary control formulation of the local rate function:

$$(5.4) \quad L_1^o(x, \beta) = \inf_{(v, \mu) \in \mathcal{A}_{x, \beta}^{1, o}} \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy),$$

where

$$\mathcal{A}_{x, \beta}^{1, o} = \left\{ v(\cdot) : \mathcal{Y} \mapsto \mathbb{R}^d, \mu \in \mathcal{P}(\mathcal{Y}) : (v, \mu) \text{ satisfy } \int_{\mathcal{Y}} \mathcal{L}_x^1 f(y) \mu(dy) = 0 \right. \\ \left. \text{for all } f \in C^2(\mathcal{Y}), \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) < \infty \text{ and } \beta = \int_{\mathcal{Y}} \lambda_1(x, y, v(y)) \mu(dy) \right\}.$$

The relaxed control formulation turns out to be more convenient when studying convergence. The fact that  $L_1^r(x, \beta) = L_1^o(x, \beta)$  follows from Jensen's inequality and that  $\lambda_1(x, y, z)$  is affine in  $z$ . To be precise, since  $(v, \mu) \in \mathcal{A}_{x, \beta}^{1, o}$  induces a  $P \in \mathcal{A}_{x, \beta}^{1, r}$  via  $P(dzdy) = \delta_{v(y)}(dz)\mu(dy)$ ,  $L_1^r(x, \beta) \leq L_1^o(x, \beta)$ . Given  $P \in \mathcal{A}_{x, \beta}^{1, r}$  we can let  $\mu$  be its  $y$ -marginal, and then define  $v(y) = \int_{\mathcal{Z}} z \eta(dz|y)$ , where  $\eta(dz|y)$  is the conditional distribution, so that  $(v, \mu) \in \mathcal{A}_{x, \beta}^{1, o}$ . By Jensen's inequality

$$\int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P(dzdy) \geq \int_{\mathcal{Y}} \frac{1}{2} \left\| \int_{\mathcal{Z}} z \eta(dz|y) \right\|^2 \mu(dy) = \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy),$$

and so  $L_1^r(x, \beta) \geq L_1^o(x, \beta)$ . The same will be true for the analogous quantities in Regimes 2 and 3, though there we will also need to use that the generator is affine in  $z$ .

An explicit expression for the local rate function (5.4) will be given in Theorem 5.2. It turns on the following technical lemma which states a Hölder inequality for integrals of matrices. The proof of the lemma is deferred to the end of this section.

**Lemma 5.1.** *Let  $\kappa \in \mathcal{L}^2(\mathcal{Y}, M_{d \times d}(\mathbb{R}); \mu)$  and  $u \in \mathcal{L}^2(\mathcal{Y}, M_{d \times 1}(\mathbb{R}); \mu)$  be matrix and vector valued functions, respectively. Define*

$$\beta = \int_{\mathcal{Y}} \kappa(y) u(y) \mu(dy) \quad \text{and} \quad q = \int_{\mathcal{Y}} \kappa(y) \kappa^T(y) \mu(dy),$$

*and assume that  $q$  is positive definite. Then*

$$\beta^T q^{-1} \beta \leq \int_{\mathcal{Y}} \|u(y)\|^2 \mu(dy).$$

**Theorem 5.2.** *Under Conditions 2.2 and 2.3, the infimization problem (5.1) and hence (5.4) has the explicit solution*

$$L_1^o(x, \beta) = \frac{1}{2} (\beta - r(x))^T q^{-1}(x) (\beta - r(x)),$$

where

$$\bullet \quad r(x) = \int_{\mathcal{Y}} (I + \frac{\partial \chi}{\partial y})(x, y) c(x, y) \mu(dy|x),$$

- $q(x) = \int_{\mathcal{Y}} (I + \frac{\partial \chi}{\partial y})(x, y) \sigma(x, y) \sigma^T(x, y) (I + \frac{\partial \chi}{\partial y})^T(x, y) \mu(dy|x),$

and where  $\mu(dy|x)$  is the unique invariant measure corresponding to the operator  $\mathcal{L}_x^1$  and  $\chi(x, y)$  is defined by (2.1). The control

$$v(y) = \bar{u}_\beta(x, y) = \sigma^T(x, y) \left( I + \frac{\partial \chi}{\partial y}(x, y) \right)^T q^{-1}(x) (\beta - r(x))$$

attains the infimum in (5.4).

*Proof.* First observe that for any  $v \in \mathcal{A}_{x, \beta}^{1, o}$

$$\int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy|x) \geq (\beta - r(x))^T q^{-1}(x) (\beta - r(x)).$$

This can be derived as follows. Any  $v \in \mathcal{A}_{x, \beta}^{1, o}$  satisfies

$$\beta = \int_{\mathcal{Y}} \lambda_1(x, y, v(y)) \mu(dy|x) = r(x) + \int_{\mathcal{Y}} \left( I + \frac{\partial \chi}{\partial y} \right) \sigma(x, y) (v(y))^T \mu(dy|x).$$

Then treating  $x$  as a parameter and applying Lemma 5.1 to the relation above with  $\beta - r(x)$  in place of  $\beta$ ,  $\kappa(x, y) = (I + \frac{\partial \chi}{\partial y}) \sigma(x, y)$  and  $u(y) = v(y)$  we immediately get the claim.

Next we observe that by choosing (with  $x$  again treated as a parameter)

$$v(y) = \bar{u}_\beta(x, y) = \sigma^T(x, y) \left( I + \frac{\partial \chi}{\partial y}(x, y) \right)^T q^{-1}(x) (\beta - r(x)),$$

we have

$$\int_{\mathcal{Y}} \|\bar{u}_\beta(x, y)\|^2 \mu(dy|x) = (\beta - r(x))^T q^{-1}(x) (\beta - r(x)).$$

This completes the proof of the theorem.  $\square$

Now we have all the ingredients to prove the Laplace principle upper bound and hence to complete the proof of the LDP for  $X^\epsilon$  in Regime 1.

*Proof of Laplace principle upper bound for Regime 1.* For each  $\epsilon > 0$ , let  $X^\epsilon$  be the unique strong solution to (1.1). To prove the Laplace principle upper bound we must show that for all bounded, continuous functions  $h$  mapping  $\mathcal{C}([0, 1]; \mathbb{R}^d)$  into  $\mathbb{R}$

$$\limsup_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \leq \inf_{\phi \in \mathcal{C}([0, 1]; \mathbb{R}^d)} [S(\phi) + h(\phi)].$$

Let  $\eta > 0$  be given and consider  $\psi \in \mathcal{C}([0, 1]; \mathbb{R}^d)$  with  $\psi_0 = x_0$  such that

$$(5.5) \quad S(\psi) + h(\psi) \leq \inf_{\phi \in \mathcal{C}([0, 1]; \mathbb{R}^d)} [S(\phi) + h(\phi)] + \eta < \infty.$$

Since  $h$  is bounded, this implies that  $S(\psi) < \infty$ , and thus  $\psi$  is absolutely continuous. Theorem 5.2 shows that  $L_1^q(x, \beta)$  is continuous and finite at each  $(x, \beta) \in \mathbb{R}^{2d}$ . By a standard mollification argument we can further assume that  $\psi$  is piecewise continuous (see for example Subsection 6.5 of [18]). Given this particular function  $\psi$  define

$$\bar{u}(t, x, y) = \sigma^T(x, y) \left( I + \frac{\partial \chi}{\partial y}(x, y) \right)^T q^{-1}(x) (\psi_t - r(x)),$$

where  $\chi$  satisfies (2.1). Clearly,  $\bar{u}(t, x, y)$  is periodic in  $y$ . Lastly, we define a control in (partial) feedback form by

$$\bar{u}^\epsilon(t) = \bar{u} \left( t, X_t^\epsilon, \frac{X_t^\epsilon}{\delta} \right).$$

Then standard homogenization theory for locally periodic diffusions and the fact that the invariant measure  $\mu(\cdot|x)$  is continuous as a function of  $x$  (see for example Chapter 3, Section 4.6 of [9]) imply the following:

(i)  $\bar{X}^\epsilon \xrightarrow{\mathcal{D}} \bar{X}$ , where w.p.1

$$\begin{aligned}
\bar{X}_t &= x_0 + \int_0^t r(\bar{X}_s) ds + \int_0^t \left[ \int_{\mathcal{Y}} \left( I + \frac{\partial \chi}{\partial y}(\bar{X}_s, y) \right) \sigma(\bar{X}_s, y) \bar{u}(s, \bar{X}_s, y) \mu(dy|\bar{X}_s) \right] ds \\
&= x_0 + \int_0^t r(\bar{X}_s) ds + \int_0^t \left[ \int_{\mathcal{Y}} \left( I + \frac{\partial \chi}{\partial y} \right) \sigma \sigma^T \left( I + \frac{\partial \chi}{\partial y} \right)^T \mu(dy|\bar{X}_s) \right] q^{-1}(\bar{X}_s) (\dot{\psi}_s - r(\bar{X}_s)) ds \\
&= x_0 + \int_0^t r(\bar{X}_s) ds + \int_0^t q(\bar{X}_s) q^{-1}(\bar{X}_s) (\dot{\psi}_s - r(\bar{X}_s)) ds \\
&= x_0 + \int_0^t \dot{\psi}_s ds \\
&= \psi_t,
\end{aligned}$$

(ii) the cost satisfies

$$(5.6) \quad \mathbb{E}_{x_0} \left( \frac{1}{2} \int_0^1 \|\bar{u}_s^\epsilon\|^2 ds - \frac{1}{2} \int_0^1 \int_{\mathcal{Y}} \|\bar{u}(s, \bar{X}_s, y)\|^2 \mu(dy|\bar{X}_s) ds \right)^2 \rightarrow 0, \text{ as } \epsilon \downarrow 0.$$

Theorem 5.2 then implies that

$$(5.7) \quad \mathbb{E}_{x_0} \int_0^1 \int_{\mathcal{Y}} \|\bar{u}(s, \bar{X}_s, y)\|^2 \mu(dy|\bar{X}_s) ds = \mathbb{E}_{x_0} S(\bar{X}) = S(\psi).$$

Thus

$$\begin{aligned}
\limsup_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] &= \limsup_{\epsilon \downarrow 0} \inf_{u \in \mathcal{A}} \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|u_t\|^2 dt + h(\bar{X}^\epsilon) \right] \\
&\leq \limsup_{\epsilon \downarrow 0} \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|\bar{u}_t^\epsilon\|^2 dt + h(\bar{X}^\epsilon) \right] \\
&\leq \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \int_{\mathcal{Y}} \|\bar{u}(s, \bar{X}_s, y)\|^2 \mu(dy|\bar{X}_s) dy ds + h(\bar{X}) \right] \\
&= [S(\psi) + h(\psi)] \\
&\leq \inf_{\phi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [S(\phi) + h(\phi)] + \eta.
\end{aligned}$$

Line 1 follows from the representation Theorem 2.4. Line 2 follows from the choice of a particular control. Line 3 follows from (5.6) and the continuity of  $h$ . Line 4 follows from (5.7) and from the fact that  $\bar{X}_t = \psi_t$ . Lastly, line 5 follows from (5.5). Since  $\eta > 0$  is arbitrary, the upper bound is proved.  $\square$

In fact, the considerations above allow us to derive an explicit representation formula for the rate function in Regime 1. We summarize the results in the following theorem.

**Theorem 5.3.** *Let  $\{X^\epsilon, \epsilon > 0\}$  be the unique strong solution to (1.1) and consider Regime 1. Under Conditions 2.2 and 2.3,  $\{X^\epsilon, \epsilon > 0\}$  satisfies a large deviations principle with rate function*

$$S(\phi) = \begin{cases} \frac{1}{2} \int_0^1 (\dot{\phi}_s - r(\phi_s))^T q^{-1}(\phi_s) (\dot{\phi}_s - r(\phi_s)) ds & \text{if } \phi \in \mathcal{C}([0,1]; \mathbb{R}^d) \text{ is absolutely continuous} \\ +\infty & \text{otherwise.} \end{cases}$$

We conclude with the proof of Lemma 5.1.

*Proof of Lemma 5.1.* Since  $q$  is positive definite and symmetric one can write

$$q^{-1} = W^T W,$$

where  $W$  is an invertible matrix. It follows that

$$\beta^T q^{-1} \beta = \|W\beta\|^2.$$

Without loss of generality we can assume

$$\int_{\mathcal{Y}} \|u(y)\|^2 \mu(dy) = 1.$$

By the Cauchy-Schwartz inequality in  $\mathbb{R}^d$  we have

$$\begin{aligned} \|W\beta\|^2 &= \left\langle W\beta, W \int_{\mathcal{Y}} \kappa(y) u(y) \mu(dy) \right\rangle \\ &= \int_{\mathcal{Y}} \langle u(y), \kappa^T(y) W^T W \beta \rangle \mu(dy) \\ &\leq \sqrt{\int_{\mathcal{Y}} \|u\|^2 \mu(dy)} \sqrt{\int_{\mathcal{Y}} \|\kappa^T(y) W^T W \beta\|^2 \mu(dy)} \\ &= \sqrt{\int_{\mathcal{Y}} \|\kappa^T(y) W^T W \beta\|^2 \mu(dy)} \\ &= \sqrt{\beta^T W^T W \left[ \int_{\mathcal{Y}} \kappa(y) \kappa^T(y) \mu(dy) \right] W^T W \beta} \\ &= \sqrt{\beta^T W^T W \beta} \\ &= \|W\beta\|. \end{aligned}$$

If  $\|W\beta\| = 0$ , then the result holds automatically. If  $\|W\beta\| \neq 0$  then we get  $\|W\beta\| \leq 1$ , which proves the result.  $\square$

**5.1. Example.** In this subsection we consider an example. A particular model of interest is the first order Langevin equation

$$(5.8) \quad dX_t^\epsilon = \left[ -\frac{\epsilon}{\delta} \nabla Q \left( \frac{X_t^\epsilon}{\delta} \right) - \nabla V(X_t^\epsilon) \right] dt + \sqrt{\epsilon} \sqrt{2D} dW_t, \quad X_0^\epsilon = x_0,$$

where  $2D$  is a diffusion constant and the two-scale potential is composed by a large-scale part,  $V(x)$ , and a fluctuating part,  $\epsilon Q(x/\delta)$ . An example of such a potential is given in Figure 1.

To connect to our notation let  $b(x, y) = -\nabla Q(y)$  and  $c(x, y) = -\nabla V(x)$ , and suppose we consider Regime 1. In this case there is an explicit formula for the invariant density  $\mu(y)$ , which is the Gibbs distribution

$$\mu(y) = \frac{1}{Z} e^{-\frac{Q(y)}{D}}, \quad Z = \int_{\mathcal{Y}} e^{-\frac{Q(y)}{D}} dy.$$

Moreover, it is easy to see that the centering Condition 2.3 holds.

When we have a separable fluctuating part, i.e.  $Q(y_1, y_2, \dots, y_d) = Q_1(y_1) + Q_2(y_2) + \dots + Q_d(y_d)$ , everything can be calculated explicitly. We summarize the results in the following corollary.



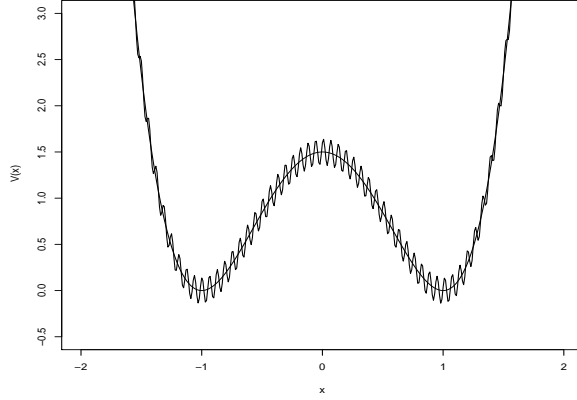


FIGURE 1.  $V^\epsilon(x, \frac{x}{\delta}) = \epsilon (\cos(\frac{x}{\delta}) + \sin(\frac{x}{\delta})) + \frac{3}{2}(x^2 - 1)^2$  and  $V(x) = \frac{3}{2}(x^2 - 1)^2$  with  $\epsilon = 0.1$  and  $\delta = 0.01$ .

**Corollary 5.4.** *Let  $\{X^\epsilon, \epsilon > 0\}$  be the unique strong solution to (5.8). Assume  $Q(y_1, y_2, \dots, y_d) = Q_1(y_1) + Q_2(y_2) + \dots + Q_d(y_d)$  and consider Regime 1. Under Condition 2.2,  $\{X^\epsilon, \epsilon > 0\}$  satisfies a large deviations principle with rate function*

$$S(\phi) = \begin{cases} \frac{1}{2} \int_0^1 (\dot{\phi}_s - r(\phi_s))^T q^{-1} (\dot{\phi}_s - r(\phi_s)) ds & \text{if } \phi \in \mathcal{C}([0, 1]; \mathbb{R}^d) \text{ is absolutely continuous} \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$r(x) = -\Theta \nabla V(x), \quad q = 2D\Theta, \quad \Theta = \text{diag} \left[ \frac{1}{Z_1 \hat{Z}_1}, \dots, \frac{1}{Z_d \hat{Z}_d} \right]$$

and for  $i = 1, 2, \dots, d$

$$Z_i = \int_{\mathbb{T}} e^{-\frac{Q_i(y_i)}{D}} dy_i, \quad \hat{Z}_i = \int_{\mathbb{T}} e^{\frac{Q_i(y_i)}{D}} dy_i.$$

Observing the effective diffusivity matrix  $q$  in Corollary 5.4, we see that the diagonal elements of  $q$  are always smaller than the corresponding diagonal elements of the original one. In the original multiscale problem there are many small energy barriers. These are not captured by the homogenized potential and hence must be accounted for in the homogenized process, and thus the trapping from the many local minima is responsible for the reduction of the diffusion coefficient.

## 6. REGIME 2: LAPLACE PRINCIPLE UPPER BOUND AND ALTERNATIVE REPRESENTATION.

In this section we prove the Laplace principle upper bound for Regime 2. We need several auxiliary results that will be proven in Subsection 6.1. For notational convenience we drop the superscript 2 from  $\bar{X}^2$  and  $P^2$ .

As was done for Regime 1 we can define the relaxed and ordinary control formulations of the local rate function,  $L_2^r(x, \beta)$  and  $L_2^o(x, \beta)$ , by considering  $\lambda_2$  and  $\mathcal{L}_{z,x}^2$  in place of  $\lambda_1$  and  $\mathcal{L}_x^1$ . For the same reasons as in Section 5 (but also using that  $\mathcal{L}_{z,x}^2$  is affine in  $z$ ), these two expressions coincide. The key difference between this case and the last is that  $\mathcal{L}_{z,x}^2$  depends on  $z$ , while  $\mathcal{L}_x^1$  did not. This means that relations between the elements of a viable pair are more complex, and in particular that the joint distribution of the control  $z$  and fast variable  $y$  is important.

Similarly to what was done in Regime 1, the limiting occupation measure  $P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y} \times [0, 1])$  can be decomposed as stochastic kernels in the form

$$P(dzdydt) = \eta(dz|y, t)\mu(dy|t)dt.$$

Moreover, by Theorem 2.8 we have  $(\bar{X}, P) \in \mathcal{V}_{\lambda_2, \mathcal{L}^2}$ . We will use that both  $\lambda_2$  and  $\mathcal{L}_{z,x}^2$  are affine in  $z$ . If  $v(t, y) : [0, 1] \times \mathcal{Y} \mapsto \mathbb{R}^d$  is defined by

$$v(t, y) = \int_{\mathcal{Z}} z\eta(dz|y, t),$$

then by viability  $\bar{X}_t$  satisfies

$$\bar{X}_t = x_0 + \int_0^t \left[ \int_{\mathcal{Y}} (\gamma b(\bar{X}_s, y) + c(\bar{X}_s, y) + \sigma(\bar{X}_s, y)v(s, y)) \mu(dy|s) \right] ds$$

where  $\mu$  is such that for all  $f \in \mathcal{C}^2(\mathcal{Y})$  and  $t \in [0, 1]$

$$\int_0^t \int_{\mathcal{Y}} \mathcal{L}_{v(s,y), \bar{X}_s}^2 f(y) \mu(dy|s) = 0.$$

*Proof of Laplace principle upper bound for Regime 2.*

We need to prove that

$$\limsup_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \leq \inf_{\phi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [S(\phi) + h(\phi)].$$

For given  $\eta > 0$  we can find  $\psi \in \mathcal{C}([0, 1]; \mathbb{R}^d)$  with  $\psi_0 = x_0$  such that

$$(6.1) \quad S(\psi) + h(\psi) \leq \inf_{\phi \in \mathcal{C}([0,1]; \mathbb{R}^d)} [S(\phi) + h(\phi)] + \eta < \infty.$$

Since  $h$  is bounded, this implies that  $S(\psi) < \infty$ , and thus  $\psi$  is absolutely continuous.

Let

$$\mathcal{A}_{x,\beta}^{2,o} = \left\{ v(\cdot) : \mathcal{Y} \mapsto \mathbb{R}^d, \mu \in \mathcal{P}(\mathcal{Y}) : (v, \mu) \text{ satisfy } \int_{\mathcal{Y}} \mathcal{L}_{v(y),x}^2 f(y) \mu(dy) = 0 \right. \\ \left. \text{for all } f \in C^2(\mathcal{Y}), \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) < \infty \text{ and } \beta = \int_{\mathcal{Y}} \lambda_2(x, y, v(y)) \mu(dy) \right\}.$$

Then the ordinary control formulation of the local rate function is

$$(6.2) \quad L_2^o(x, \beta) = \inf_{(v,\mu) \in \mathcal{A}_{x,\beta}^{2,o}} \left\{ \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) \right\}.$$

Calling this an “ordinary control formulation” is perhaps a bit misleading. Invariant measures are in general characterized by equations of the form

$$(6.3) \quad \int_{\mathcal{Y}} \mathcal{L}_{v(y),x}^2 f(y) \mu(dy) = 0$$

where  $v(\cdot) : \mathcal{Y} \mapsto \mathbb{R}^d$  plays the role of a feedback control. In the definition of  $\mathcal{A}_{x,\beta}^{2,o}$  no claim is made that  $\mu$  is an invariant distribution for any controlled dynamics. [This was not an issue in Regime 1 since  $\mathcal{L}_x^1$  did not depend on  $z$ . Hence there was only one invariant distribution that did not depend in any way on the control.] In fact for some choices of  $v$  it may be difficult to argue that an invariant distribution corresponding to  $\mathcal{L}_{v(y),x}^2$  exists. However, we will use results from [29] that allow us to represent  $L_2^o(x, \beta)$  in terms of the average cost of an ergodic control problem for which the Bellman equation has a classical sense solution. This will lead to a control  $v$  that is bounded and Lipschitz continuous, and hence for the corresponding controlled diffusion there will be a unique invariant distribution  $\mu$  such that the pair satisfy (6.3).

By Theorem 6.3 below,  $L_2^o(x, \beta)$  is continuous and finite at each  $(x, \beta) \in \mathbb{R}^{2d}$ . Thus, by a standard mollification argument, we can further assume that  $\dot{\psi}$  is piecewise constant (see for example Subsection 6.5 in [18]). Theorem 6.2 below implies that there is  $\bar{u}(t, x, y)$  that is bounded, continuous in  $x$  and Lipschitz continuous  $y$ , and piecewise constant in  $t$  and which satisfies

$$(6.4) \quad \bar{u}(t, x, \cdot) \in \operatorname{argmin}_v \left\{ \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) : (v, \mu) \in \mathcal{A}_{x, \psi_t}^{2,o} \right\}.$$

As remarked previously for this particular control  $\bar{u}$  the invariant measure corresponding to the operator  $\mathcal{L}_{\bar{u}, x}^2$  is unique and will be denoted by  $\bar{\mu}_{\bar{u}}(dy)$ . The control used in the large deviation problem (in feedback form) is then

$$\bar{u}^\epsilon(t) = \bar{u} \left( t, \bar{X}_t^\epsilon, \frac{\bar{X}_t^\epsilon}{\delta} \right).$$

Since  $\sigma\sigma^T$  is uniformly nondegenerate and Lipschitz continuous and since  $\bar{u}$  is continuous in  $x$  and  $y$ , a strong solution to (2.2) exists. By standard averaging theory and the fact that  $\bar{\mu}_{\bar{u}(t, x, \cdot)}(\cdot)$  is continuous in  $x$  (Theorem 6.2) and piecewise continuous in  $t$  we have that  $\bar{X}^\epsilon \xrightarrow{\mathcal{D}} \bar{X}$ , where

$$\bar{X}_t = x_0 + \int_0^t \int_{\mathcal{Y}} \lambda_2(\bar{X}_s, y, \bar{u}(s, \bar{X}_s, y)) \bar{\mu}_{\bar{u}(s, \bar{X}_s, \cdot)}(dy) ds.$$

Since (6.4) holds we get, for  $\psi$  such that  $\psi_0 = x_0$ ,

$$\bar{X}_t = x_0 + \int_0^t \dot{\psi}_s ds = \psi_t \quad \text{for any } t \in [0, 1], \text{ w.p.1.}$$

Taking into account the above facts, we have the following chain of inequalities:

$$\begin{aligned} \limsup_{\epsilon \downarrow 0} \left[ -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \right] &= \limsup_{\epsilon \downarrow 0} \inf_u \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|u_t\|^2 dt + h(\bar{X}^\epsilon) \right] \\ &\leq \limsup_{\epsilon \downarrow 0} \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \|\bar{u}^\epsilon(t)\|^2 dt + h(\bar{X}^\epsilon) \right] \\ &= \mathbb{E}_{x_0} \left[ \frac{1}{2} \int_0^1 \int_{\mathcal{Y}} \|\bar{u}(t, \bar{X}_t, y)\|^2 \bar{\mu}_{\bar{u}(t, \bar{X}_t, \cdot)}(dy) dt + h(\bar{X}) \right] \\ &= \mathbb{E}_{x_0} [S(\bar{X}) + h(\bar{X})] \\ &= S(\psi) + h(\psi) \\ &\leq \inf_{\phi \in \mathcal{C}([0, 1]; \mathbb{R}^d)} [S(\phi) + h(\phi)] + \eta. \end{aligned}$$

Line 1 follows from the representation Theorem 2.4. Line 2 follows from the choice of the particular control. Line 3 follows from the definition of the control by the minimization problem above and the continuity of  $h$ . Line 4 follows from the definition of  $S$ . Line 6 is from (6.1). Finally, since  $\eta$  is arbitrary, we are done.  $\square$

In fact, the considerations above allow us to derive an alternative representation formula for the rate function in Regime 2. We summarize the results in the following theorem.

**Theorem 6.1.** *Let  $\{X^\epsilon, \epsilon > 0\}$  be the unique strong solution to (1.1) such that Condition 2.2 holds and assume that we are considering Regime 2. Then  $\{X^\epsilon, \epsilon > 0\}$  satisfies a large deviations principle with rate function*

$$S(\phi) = \begin{cases} \int_0^1 L_2^o(\phi_s, \dot{\phi}_s) ds & \text{if } \phi \in \mathcal{C}([0, 1]; \mathbb{R}^d) \text{ is absolutely continuous} \\ +\infty & \text{otherwise.} \end{cases}$$

**6.1. Properties of the local rate function and of the optimal control for Regime 2.** In this section we study the local rate function  $L_2^o(x, \beta) = \inf_{(v, \mu) \in \mathcal{A}_{x, \beta}^{2, o}} \left\{ \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) \right\}$ . The main theorems of this section are the following two.

**Theorem 6.2.** *Assume Condition 2.2. Then there is a pair  $(\bar{u}, \bar{\mu})$  that achieves the infimum in the definition of the local rate function such that  $\bar{u} = \bar{u}_\beta(x, y)$  is, for each fixed  $\beta \in \mathbb{R}^d$ , continuous in  $x$ , Lipschitz continuous in  $y$  and measurable in  $(x, y, \beta)$ . Moreover,  $\bar{\mu}(dy) = \bar{\mu}_{\bar{u}}(dy|x)$  is the unique invariant measure corresponding to the operator  $\mathcal{L}_{\bar{u}_\beta(x, y), x}^2$  and it is weakly continuous as a function of  $x$ .*

**Theorem 6.3.** *Assume Condition 2.2. Then, the local rate function  $L_2^o(x, \beta)$  is finite, continuous at each  $(x, \beta) \in \mathbb{R}^{2d}$  and differentiable with respect to  $\beta$ .*

The proof of these theorems will be given in several steps. In Lemma 6.4 we prove that  $L_2^o$  is convex in  $\beta$  and finite. One of the consequences of this lemma is that the subdifferential of  $L_2^o(x, \cdot)$  is non empty. This result is used by Lemma 6.5 where we rewrite  $L_2^o$  in the spirit of a Lagrange multiplier problem where the role of the Lagrange multiplier is played by an element in the subdifferential of  $L_2^o(x, \cdot)$ . Then, using Lemma 6.5 we prove in Lemma 6.6 that an optimal control exists which is bounded and Lipschitz continuous in  $y$ . Lemma 6.8 uses Lemmas 6.4 and 6.6 together with the technical Lemma 6.7 to prove that the dual of  $L_2^o(x, \beta)$  with respect to  $\beta$  is strictly convex, which implies that  $L_2^o(x, \beta)$  is differentiable in  $\beta$ . In Lemma 6.9 we prove that  $L_2^o(x, \beta)$  is continuous in  $(x, \beta) \in \mathbb{R}^d$  using Lemmas 6.4 and 6.6. Lastly, in Lemma 6.10 we prove that the control that is constructed in the proof of Lemma 6.6 is continuous in  $x$ , which together with uniqueness of the corresponding invariant measure imply that the latter is weakly continuous in  $x$ . Theorem 6.2 follows from Lemmas 6.6 and 6.10. Theorem 6.3 follows from Lemmas 6.4, 6.8 and 6.9.

For the reader's convenience we recall

$$\mathcal{A}_{x, \beta}^{2, r} = \left\{ P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y}) : \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_{z, x}^2 f(y) P(dz dy) = 0 \text{ for all } f \in C^2(\mathcal{Y}) \right. \\ \left. \int_{\mathcal{Z} \times \mathcal{Y}} \|z\|^2 P(dz dy) < \infty \text{ and } \beta = \int_{\mathcal{Z} \times \mathcal{Y}} \lambda_2(x, y, z) P(dz dy) \right\}.$$

For notational convenience, we ignore for the moment the  $x$ -dependence since this is seen as parameter by the local rate function. Sometimes, the analysis works with the relaxed form of the local rate, but as noted previously  $L_2^o(\beta) = L_2^r(\beta)$ .

**Lemma 6.4.** *The cost  $L_2^r(\beta)$  is a finite and convex function of  $\beta$ .*

*Proof.* Given  $\beta$  let  $v_\beta(y) = \sigma^{-1}(y)(\beta - \gamma b(y) - c(y))$ . Then  $v_\beta(y)$  is Lipschitz continuous, and hence there is an associated unique invariant distribution  $\mu_\beta(dy)$ . Letting  $P(dz dy) = \delta_{v_\beta(y)}(dz) \mu_\beta(dy)$ , we have

$$\int_{\mathcal{Z} \times \mathcal{Y}} (\gamma b(y) + c(y) + \sigma(y)z) P(dz dy) = \int_{\mathcal{Y}} \beta \mu_\beta(dy) = \beta,$$

and similarly the first condition for inclusion in  $\mathcal{A}_\beta^{2, r}$  can be checked. Since  $v_\beta(y)$  is bounded the associated cost is finite, and so  $L_2^r(\beta) < \infty$ .

Next let  $\beta_1, \beta_2 \in \mathbb{R}^d$  and denote by  $P_1, P_2$  corresponding controls such that  $\int_{\mathcal{Z} \times \mathcal{Y}} \lambda(y, z) P_i(dzdy) = \beta_i$ . Consider a parameter  $\eta \in [0, 1]$  and define  $P_0 = \eta P_1 + (1 - \eta) P_2$ . Due to the linearity of integration,  $P_0 \in A_{\eta\beta_1 + (1-\eta)\beta_2}^{2,r}$ , and therefore

$$\begin{aligned} L_2^r(\eta\beta_1 + (1 - \eta)\beta_2) &\leq \int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P_0(dzdy) \\ &= \eta \int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P_1(dzdy) + (1 - \eta) \int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P_2(dzdy). \end{aligned}$$

Taking the infimum over all admissible  $P_1, P_2$  we get

$$L_2^r(\eta\beta_1 + (1 - \eta)\beta_2) \leq \eta L_2^r(\beta_1) + (1 - \eta) L_2^r(\beta_2).$$

This proves the convexity, and completes the proof of the lemma.  $\square$

For any  $\beta \in \mathbb{R}^d$  the subdifferential of  $L_2^r$  at  $\beta$  is defined by

$$\partial L_2^r(\beta) = \{\zeta \in \mathbb{R}^d : L_2^r(\beta') - L_2^r(\beta) \geq \zeta \cdot (\beta' - \beta) \text{ for all } \beta' \in \mathbb{R}^d\}.$$

Since  $L_2^r$  is finite and convex  $\partial L_2^r(\beta)$  is always nonempty. Define

$$\begin{aligned} \mathcal{B}^{2,r} &= \left\{ P \in \mathcal{P}(\mathcal{Z} \times \mathcal{Y}) : \int_{\mathcal{Z} \times \mathcal{Y}} \|z\|^2 P(dzdy) < \infty, \int_{\mathcal{Z} \times \mathcal{Y}} \mathcal{L}_z^2 f(y) P(dzdy) = 0 \text{ for all } f \in C^2(\mathcal{Y}) \right\} \\ \mathcal{B}^{2,o} &= \left\{ v(\cdot) : \mathcal{Y} \mapsto \mathbb{R}^d, \mu \in \mathcal{P}(\mathcal{Y}) : \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) < \infty, \int_{\mathcal{Y}} \mathcal{L}_{v(y)}^2 f(y) \mu(dy) = 0 \text{ for all } f \in C^2(\mathcal{Y}) \right\} \end{aligned}$$

and for  $\zeta \in \mathbb{R}^d$  let

$$\tilde{L}_2^r(\zeta) = \inf_{P \in \mathcal{B}^{2,r}} \int_{\mathcal{Z} \times \mathcal{Y}} \left( \frac{1}{2} \|z\|^2 - \zeta \cdot (\gamma b(y) + c(y) + \sigma(y)z) \right) P(dzdy).$$

We have the following lemma.

**Lemma 6.5.** *Consider any  $\beta \in \mathbb{R}^d$  and any  $\zeta_\beta \in \partial L_2^r(\beta)$ . Then*

$$\tilde{L}_2^r(\zeta_\beta) = L_2^r(\beta) - \zeta_\beta \cdot \beta.$$

*Proof.* First we prove that  $\tilde{L}_2^r(\zeta_\beta) \leq L_2^r(\beta) - \zeta_\beta \cdot \beta$ , which follows from

$$\begin{aligned} L_2^r(\beta) - \zeta_\beta \cdot \beta &= \inf_{P \in \mathcal{A}_\beta^{2,r}} \int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P(dzdy) - \zeta_\beta \cdot \beta \\ &= \inf_{P \in \mathcal{A}_\beta^{2,r}} \int_{\mathcal{Z} \times \mathcal{Y}} \left( \frac{1}{2} \|z\|^2 - \zeta_\beta \cdot (\gamma b(y) + c(y) + \sigma(y)z) \right) P(dzdy) \\ &\geq \inf_{P \in \mathcal{B}^{2,r}} \int_{\mathcal{Z} \times \mathcal{Y}} \left( \frac{1}{2} \|z\|^2 - \zeta_\beta \cdot (\gamma b(y) + c(y) + \sigma(y)z) \right) P(dzdy) \\ &= \tilde{L}_2^r(\zeta_\beta). \end{aligned}$$

For the opposite direction we use that  $\zeta_\beta \in \partial L_2^r(\beta)$ . Consider any  $\beta' \in \mathbb{R}^d$  and any  $P \in \mathcal{A}_{\beta'}^{2,r}$ . Then

$$\begin{aligned} L_2^r(\beta) - \zeta_\beta \cdot \beta &\leq L_2^r(\beta') - \zeta_\beta \cdot \beta' \\ &\leq \int_{\mathcal{Z} \times \mathcal{Y}} \frac{1}{2} \|z\|^2 P(dzdy) - \zeta_\beta \cdot \beta' \\ &= \int_{\mathcal{Z} \times \mathcal{Y}} \left( \frac{1}{2} \|z\|^2 - \zeta_\beta \cdot (\gamma b(y) + c(y) + \sigma(y)z) \right) P(dzdy). \end{aligned}$$

Since  $\mathcal{B}^{2,r} = \cup_{\beta' \in \mathbb{R}^d} A_{\beta'}^{2,r}$ , the last display implies

$$L_2^r(\beta) - \zeta_\beta \cdot \beta \leq \inf_{P \in \mathcal{B}^{2,r}} \int_{\mathcal{Z} \times \mathcal{Y}} \left( \frac{1}{2} \|z\|^2 - \zeta_\beta \cdot (\gamma b(y) + c(y) + \sigma(y)z) \right) P(dz dy) = \tilde{L}_2^r(\zeta_\beta).$$

This concludes the proof of the lemma.  $\square$

**Lemma 6.6.** *Assume Condition 2.2. Then there is a pair  $(\bar{u}, \bar{\mu})$  that achieves the infimum in the definition of the local rate function such that  $\bar{u} = \bar{u}_\beta(x, y)$  is, for any fixed  $(x, \beta) \in \mathbb{R}^{2d}$ , bounded and Lipschitz continuous in  $y$ . Also,  $\bar{\mu}(dy) = \bar{\mu}_{\bar{u}}(dy|x)$  is the unique invariant measure corresponding to the operator  $\mathcal{L}_{\bar{u}_\beta(x,y),x}^2$ .*

*Proof.* By Lemma 6.5 we get that for any  $\beta \in \mathbb{R}^d$  and  $\zeta_\beta \in \partial L_2^r(\beta)$ ,

$$L_2^r(\beta) = \tilde{L}_2^r(\zeta_\beta) + \zeta_\beta \cdot \beta = \inf_{P \in \mathcal{B}^{2,r}} \int_{\mathcal{Z} \times \mathcal{Y}} \left( \frac{1}{2} \|z\|^2 - \zeta_\beta \cdot (\gamma b(y) + c(y) + \sigma(y)z - \beta) \right) P(dz dy).$$

According to Theorem 6.1 in [29], this optimization also has a representation via an ergodic control problem of the form

$$\tilde{L}_2^r(\zeta_\beta) + \zeta_\beta \cdot \beta = \inf \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \int_0^T \left( \frac{1}{2} \|v_s\|^2 - \zeta_\beta \cdot (\gamma b(Y_s) + c(Y_s) + \sigma(Y_s)v_s - \beta) \right) ds,$$

where the infimum is over all progressively measurable controls  $v$  and solutions to the controlled martingale problem associated with  $\mathcal{L}_z^2$ . [The paper [29] works with relaxed controls, but since here the dynamics are affine in the control and the cost is convex, the infima over relaxed and ordinary controls are the same.]

The Bellman equation associated with this control problem is

$$(6.5) \quad \inf_v \left[ \mathcal{L}_v^2 W(y) + \frac{1}{2} \|v\|^2 - \zeta_\beta \cdot (\gamma b(y) + c(y) + \sigma(y)v - \beta) \right] = \rho.$$

Using the standard vanishing discount approach and taking into account the periodicity condition (see for example [2, 7]) one can show that there is a unique pair  $(W, \rho) \in \mathcal{C}^2(\mathbb{R}^d) \times \mathbb{R}$ , such that  $W(0) = 0$  and  $W(y)$  is periodic in  $y$  with period 1 that satisfies (6.5). Since we have a classical sense solution, by the verification theorem for ergodic control  $\rho = \rho(\beta) = L_2^r(\beta) = \tilde{L}_2^r(\zeta_\beta) + \zeta_\beta \cdot \beta$ . In order to emphasize the dependence of  $W(\cdot)$  on  $\beta$  we write  $W(y) = W_\beta(y)$ . It also follows from the verification argument that an optimal control  $\bar{u}_\beta(y)$  is given by  $\bar{u}_\beta(y) = -\sigma(y)^T (\nabla_y W_\beta(y) - \zeta_\beta)$ . Compactness of the state space and the assumptions on the coefficients guarantee that the gradient of  $W_\beta(\cdot)$  is bounded, i.e.,  $\|\nabla_y W_\beta\| \leq K(\beta)$  for some constant  $K(\beta)$  that may depend on  $\beta$ . Therefore, such an optimal control is indeed bounded and Lipschitz continuous in  $y$ . Existence and uniqueness of the invariant measure follows from the latter and the non-degeneracy assumption.  $\square$

Next, we prove that the local rate function  $L_2^g(\beta)$  is actually differentiable in  $\beta \in \mathbb{R}^d$ . Recall the operator

$$\mathcal{L}_{u(y)}^2 = [\gamma b(y) + c(y) + \sigma(y)u(y)] \cdot \nabla_y + \gamma \frac{1}{2} \sigma(y) \sigma(y)^T : \nabla_y \nabla_y.$$

For notational convenience we omit the superscript 2 and write  $\mathcal{L}_{u(y)}$  in place of  $\mathcal{L}_{u(y)}^2$ . Recall also that for a bounded and Lipschitz continuous control  $\bar{u}$  there exists a unique invariant measure  $\mu(dy)$  corresponding to  $\mathcal{L}_{\bar{u}(y)}$ .

Define the set of functions

$$\mathcal{H} \doteq \left\{ h : \mathcal{Y} \mapsto \mathbb{R} \text{ such that } h \text{ is periodic, bounded, Lipschitz continuous and } \int_{\mathcal{Y}} h(y) \mu(dy) = 1 \right\}.$$



For a vector  $\theta \in \mathbb{R}^d$ ,  $\eta \in \mathbb{R}$  and  $h \in \mathcal{H}$  define the perturbed control

$$(6.6) \quad \bar{u}_\eta(y) \doteq \bar{u}(y) + \eta \sigma(y)^{-1} \theta h(y).$$

For each  $\eta$  there is a unique invariant measure  $\mu_\eta(dy)$  corresponding to  $\mathcal{L}_\eta = \mathcal{L}_{\bar{u}_\eta(y)}$ , and it is straightforward to show that  $\mu_\eta(dy) \rightarrow \mu(dy)$  in the weak topology as  $|\eta| \downarrow 0$ . Moreover, under Condition 2.2, Lemma 3.2 in [17] guarantees that the invariant measures  $\mu_\eta(dy)$  and  $\mu(dy)$  have densities  $m_\eta(y)$  and  $m(y)$  respectively. In particular, there exist unique weak sense solutions to the equations

$$\mathcal{L}_\eta^* m_\eta(y) = 0, \quad \int_{\mathcal{Y}} m_\eta(y) dy = 1 \quad \text{and} \quad \mathcal{L}_{\bar{u}}^* m(y) = 0, \quad \int_{\mathcal{Y}} m(y) dy = 1$$

where  $\mathcal{L}_\eta^*$  and  $\mathcal{L}_{\bar{u}}^*$  are the formal adjoint operators to  $\mathcal{L}_\eta$  and  $\mathcal{L}_{\bar{u}}$  respectively. The densities are strictly positive, continuous and in  $H^1(\mathcal{Y})$ . Observe that

$$(6.7) \quad \mathcal{L}_\eta^* m_\eta(y) = 0 \Leftrightarrow \mathcal{L}_{\bar{u}}^* m_\eta(y) = \eta \theta \cdot \nabla (h(y) m_\eta(y)).$$

in the weak sense.

Next, for  $g \in L^2(\mathcal{Y})$  consider the auxiliary partial differential equation

$$(6.8) \quad \mathcal{L}_{\bar{u}} f(y) = g(y) - \int_{\mathcal{Y}} g(y) \mu(dy), \quad f \text{ is 1 periodic and } \int_{\mathcal{Y}} f(y) \mu(dy) = 0.$$

By the Fredholm alternative and the strong maximum principle this equation has a unique solution. Standard elliptic regularity theory yields  $f \in H^2(\mathbb{R}^d)$ . Then by Sobolev's embedding lemma we have that  $f \in C^1(\mathbb{R}^d)$ .

Denote by  $(\cdot, \cdot)_2$  the usual inner product in  $L^2(\mathcal{Y})$ . The following lemma will be useful in the sequel.

**Lemma 6.7.** *Let  $g \in L^2(\mathcal{Y})$ ,  $\eta \in \mathbb{R}$ ,  $h \in \mathcal{H}$  and  $f \in H^2(\mathbb{R}^d)$  the solution to (6.8). Then,*

$$(g, (m_\eta - m))_2 = -\eta (\theta \cdot \nabla f, hm)_2 - \eta (\theta \cdot \nabla f, h(m_\eta - m))_2$$

*Proof.* Keeping in mind (6.7) and that  $m_\eta(y)$  and  $m(y)$  are densities, the following hold

$$\begin{aligned} (f, \mathcal{L}_{\bar{u}}^* (m_\eta - m))_2 &= \eta (f, \theta \cdot \nabla (hm_\eta))_2 \Rightarrow \\ (\mathcal{L}_{\bar{u}} f, (m_\eta - m))_2 &= -\eta (\theta \cdot \nabla f, hm_\eta)_2 \Rightarrow \\ (g, (m_\eta - m))_2 &= -\eta (\theta \cdot \nabla f, hm)_2 - \eta (\theta \cdot \nabla f, h(m_\eta - m))_2. \end{aligned}$$

This concludes the proof of the lemma.  $\square$

By Lemma 6.4 we already know that  $L_2^o$  is finite and convex. To show that  $L_2^o$  is differentiable, it is enough to show its Legendre transform is strictly convex. For  $\alpha \in \mathbb{R}^d$  define

$$\begin{aligned} H(\alpha) &\doteq \sup_{\beta \in \mathbb{R}^d} [\langle \alpha, \beta \rangle - L_2^o(\beta)] \\ (6.9) \quad &= \sup_{(v, \mu) \in \mathcal{B}^{2,o}} \left[ \left\langle \alpha, \int_{\mathcal{Y}} (\gamma b(y) + c(y) + \sigma(y)v(y)) \mu(dy) \right\rangle - \int_{\mathcal{Y}} \frac{1}{2} \|v(y)\|^2 \mu(dy) \right]. \end{aligned}$$

**Lemma 6.8.** *The Legendre transform  $H$  of  $L_2^o$  is a strictly convex function of  $\alpha \in \mathbb{R}^d$ .*

*Proof.* Suppose that  $H$  is not strictly convex. Then there are  $\alpha_i \in \mathbb{R}^d, i = 1, 2$  not equal such that for all  $\xi \in [0, 1]$

$$\begin{aligned} H(\xi \alpha_1 + (1 - \xi) \alpha_2) &= \xi H(\alpha_1) + (1 - \xi) H(\alpha_2) \\ &= \left\langle \xi \alpha_1 + (1 - \xi) \alpha_2, \int_{\mathcal{Y}} (\gamma b(y) + c(y) + \sigma(y) \bar{u}(y)) \mu(dy) \right\rangle - \int_{\mathcal{Y}} \frac{1}{2} \|\bar{u}\|^2 \mu(dy), \end{aligned}$$

where  $\bar{\beta} \doteq \int_{\mathcal{Y}} (\gamma b(y) + c(y) + \sigma(y)\bar{u}(y)) \mu(dy) \in \partial H(\xi\alpha_1 + (1-\xi)\alpha_2)$  for all  $\xi \in [0, 1]$ . As in Lemma 6.6, it can be shown that  $\bar{u}$  exists and can be chosen to be bounded and Lipschitz continuous. Also,  $\mu$  is the unique invariant measure corresponding to the operator  $\mathcal{L}_{\bar{u}(y)}$ . We will argue that the last display is impossible.

First observe that by subtracting  $\langle \alpha, \bar{\beta} \rangle$  we can arrange that  $H$  is constant for  $\alpha = \xi\alpha_1 + (1-\xi)\alpha_2$ ,  $\xi \in [0, 1]$ . Let

$$\bar{H}(\alpha) = H(\alpha) - \langle \alpha, \bar{\beta} \rangle = - \int_{\mathcal{Y}} \frac{1}{2} \|\bar{u}\|^2 \mu(dy).$$

Consider  $\xi = 1/2 + \eta$  with  $\eta$  small (and possibly negative). We will construct  $(v, \mu) \in \mathcal{B}^{2,o}$  that will give a lower bound for  $H(\xi\alpha_1 + (1-\xi)\alpha_2)$  through (6.9) that is strictly bigger than  $-\int_{\mathcal{Y}} \frac{1}{2} \|\bar{u}\|^2 \mu(dy)$ . This contradicts the constancy of  $H(\xi\alpha_1 + (1-\xi)\alpha_2)$  for  $\xi \in [0, 1]$ , and thus implies that  $H$  is strictly convex.

Define  $\bar{u}_\eta(y)$  by (6.6) with  $\theta \doteq \alpha_1 - \alpha_2$  and  $h \in \mathcal{H}$ . For  $\xi = 1/2 + \eta$  we have  $\alpha = \frac{1}{2}(\alpha_1 + \alpha_2) + \eta(\alpha_1 - \alpha_2)$ . The definition of  $H(\alpha)$  by (6.9) implies

$$\begin{aligned} \bar{H}(\alpha) &\geq \left\langle \alpha, \eta(\alpha_1 - \alpha_2) \int_{\mathcal{Y}} h(y) m_\eta(y) dy \right\rangle + \int_{\mathcal{Y}} \langle \alpha, (\gamma b(y) + c(y) + \sigma(y)\bar{u}(y)) \rangle (m_\eta(y) - m(y)) dy \\ &\quad - \int_{\mathcal{Y}} \frac{1}{2} \|\bar{u}(y) + \eta\sigma^{-1}(y)(\alpha_1 - \alpha_2)h(y)\|^2 m_\eta(y) dy. \end{aligned}$$

For  $i = 1, \dots, d$ , let  $\phi_i(y)$  be the solution to (6.8) with  $g(y) = q_i(y)$ , the  $i^{\text{th}}$  component of  $q(y) = \gamma b(y) + c(y) + \sigma(y)\bar{u}(y)$ . We write  $\phi = (\phi_1, \dots, \phi_d)$ , and also denote by  $\psi(y)$  the solution to (6.8) with  $g(y) = \|\bar{u}(y)\|^2$ . Then by Lemma 6.7 the last display can be rewritten as

$$\begin{aligned} \bar{H}(\alpha) &\geq \frac{1}{2}\eta \left[ \langle \alpha_1 + \alpha_2, (\alpha_1 - \alpha_2) \rangle - \int_{\mathcal{Y}} \left\langle \alpha_1 + \alpha_2, \frac{\partial \phi}{\partial y}(y)(\alpha_1 - \alpha_2) \right\rangle h(y) m(y) dy \right. \\ &\quad \left. - 2 \int_{\mathcal{Y}} \langle \bar{u}(y), \sigma^{-1}(y)(\alpha_1 - \alpha_2) \rangle h(y) m(y) dy \right. \\ &\quad \left. + \int_{\mathcal{Y}} \langle \alpha_1 - \alpha_2, \nabla \psi(y) \rangle h(y) m(y) dy \right] - \frac{1}{2} \int_{\mathcal{Y}} \|\bar{u}(y)\|^2 m(y) dy + o(\eta) \end{aligned}$$

where  $o(\eta)$  is such that  $o(\eta)/\eta \downarrow 0$  as  $|\eta| \downarrow 0$  and can be neglected.

Now for small  $\eta$  (perhaps negative) this is strictly bigger than  $-\frac{1}{2} \int_{\mathcal{Y}} \|\bar{u}\|^2 \mu(dy)$  unless the  $O(\eta)$  term is zero, i.e., unless

$$\begin{aligned} 0 &= \langle \alpha_1 + \alpha_2, (\alpha_1 - \alpha_2) \rangle - \int_{\mathcal{Y}} \left\langle \alpha_1 + \alpha_2, \frac{\partial \phi}{\partial y}(y)(\alpha_1 - \alpha_2) \right\rangle h(y) m(y) dy \\ &\quad - 2 \int_{\mathcal{Y}} \langle \bar{u}(y), \sigma^{-1}(y)(\alpha_1 - \alpha_2) \rangle h(y) m(y) dy \\ &\quad + \int_{\mathcal{Y}} \langle \alpha_1 - \alpha_2, \nabla \psi(y) \rangle h(y) m(y) dy. \end{aligned}$$

However, in the argument by contradiction  $\alpha_1$  and  $\alpha_2$  can be replaced by any  $\varepsilon_1\alpha_1 + (1-\varepsilon_1)\alpha_2$  and  $\varepsilon_2\alpha_1 + (1-\varepsilon_2)\alpha_2$ , so long as  $0 \leq \varepsilon_1 < \varepsilon_2 \leq 1$ . After performing this substitution and some

algebra, the last display becomes

$$\begin{aligned}
0 &= (\epsilon_1^2 - \epsilon_2^2) \int_{\mathcal{Y}} \left\langle \alpha_1 - \alpha_2, \left( I - \frac{\partial \phi}{\partial y}(y) \right) (\alpha_1 - \alpha_2) \right\rangle h(y) m(y) dy \\
&+ 2(\epsilon_1 - \epsilon_2) \left[ \int_{\mathcal{Y}} \left\langle \alpha_2, \left( I - \frac{\partial \phi}{\partial y}(y) \right) (\alpha_1 - \alpha_2) \right\rangle h(y) m(y) dy \right. \\
&\quad \left. - \int_{\mathcal{Y}} \langle \bar{u}(y), \sigma^{-1}(y)(\alpha_1 - \alpha_2) \rangle h(y) m(y) dy + \frac{1}{2} \int_{\mathcal{Y}} \langle \alpha_1 - \alpha_2, \nabla \psi(y) \rangle h(y) m(y) dy \right].
\end{aligned}$$

We claim that the last display cannot be true since  $\epsilon_1 \neq \epsilon_2$  and  $\alpha_1 - \alpha_2 \neq 0$ . By considering various choices for  $\epsilon_1$  and  $\epsilon_2$ , it is enough show that the term multiplying  $(\epsilon_1^2 - \epsilon_2^2)$  is not zero for all  $h \in \mathcal{H}$ . Let us assume the contrary, and that for all  $h \in \mathcal{H}$

$$(6.10) \quad \int_{\mathcal{Y}} \left\langle \alpha_1 - \alpha_2, \left( I - \frac{\partial \phi}{\partial y}(y) \right) (\alpha_1 - \alpha_2) \right\rangle h(y) m(y) dy = 0.$$

This implies that

$$(6.11) \quad \left\langle \alpha_1 - \alpha_2, \frac{\partial \phi}{\partial y}(y)(\alpha_1 - \alpha_2) \right\rangle = \|\alpha_1 - \alpha_2\|^2 \text{ for all } y \in \mathcal{Y}.$$

Define

$$\Phi(y) \doteq (\alpha_1 - \alpha_2) \cdot \phi(y)$$

Then  $\Phi$  is a periodic, bounded and  $C^1(\mathbb{R}^d)$  function. Consider any trajectory  $\zeta_t : \mathbb{R}_+ \mapsto \mathcal{Y}$  such that  $\dot{\zeta}_t = (\alpha_1 - \alpha_2)$ . Differentiation of  $\Phi(\zeta_t)$  and use of (6.11) give

$$\frac{d}{dt} \Phi(\zeta_t) = \left\langle \alpha_1 - \alpha_2, \frac{\partial \phi}{\partial y}(\zeta_t)(\alpha_1 - \alpha_2) \right\rangle = \|\alpha_1 - \alpha_2\|^2 > 0,$$

which cannot be true due to the periodicity and boundedness of  $\Phi$ . This implies that (6.10) is false, i.e., that there is  $h \in \mathcal{H}$  such that

$$\int_{\mathcal{Y}} \left\langle \alpha_1 - \alpha_2, \left( I - \frac{\partial \phi}{\partial y}(y) \right) (\alpha_1 - \alpha_2) \right\rangle h(y) m(y) dy \neq 0.$$

This concludes the proof of the lemma.  $\square$

Let us now recall the  $x$ -dependence and prove that the local rate function  $L_2^o(x, \beta)$  is continuous in  $(x, \beta) \in \mathbb{R}^{2d}$ .

**Lemma 6.9.** *The local rate function  $L_2^o(x, \beta)$  is continuous in  $(x, \beta) \in \mathbb{R}^{2d}$ .*

*Proof.* First, we prove that  $L_2^o(x, \beta)$  is lower semicontinuous in  $(x, \beta) \in \mathbb{R}^{2d}$ . We work with the relaxed formulation of the local rate function, but as noted previously  $L_2^r(x, \beta) = L_2^o(x, \beta)$ .

Consider  $\{(x_n, \beta_n) \in \mathbb{R}^{2d}\}_{n \in \mathbb{N}}$  such that  $(x_n, \beta_n) \rightarrow (x, \beta)$ . We want to prove

$$\liminf_{n \rightarrow \infty} L_2^r(x_n, \beta_n) \geq L_2^r(x, \beta).$$

Let  $M < \infty$  such that  $\liminf_{n \rightarrow \infty} L_2^r(x_n, \beta_n) \leq M$ . The definition of  $L_2^r(x_n, \beta_n)$  implies that we can find measures  $\{P^n, n < \infty\}$  satisfying  $P^n \in \mathcal{A}_{x_n, \beta_n}^{2,r}$  such that

$$(6.12) \quad \sup_{n < \infty} \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y}} \|z\|^2 P^n(dz dy) < M + 1$$

and

$$L_2^r(x_n, \beta_n) \geq \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y}} \|z\|^2 P^n(dz dy) - \frac{1}{n} \right]$$

It follows from (6.12) and the definition of  $\mathcal{A}_{x,\beta}^{2,r}$  that  $\{P^n, n < \infty\}$  is tight and any limit point  $P$  of  $P^n$  will be in  $\mathcal{A}_{x,\beta}^{2,r}$ . Hence by Fatou's Lemma

$$\begin{aligned} \liminf_{n \rightarrow \infty} L_2^r(x_n, \beta_n) &\geq \liminf_{n \rightarrow \infty} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y}} \|z\|^2 P^n(dz dy) - \frac{1}{n} \right] \\ &\geq \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times \mathcal{X}} \|z\|^2 P(dz dy) \\ &\geq \inf_{P \in \mathcal{A}_{x,\beta}^{2,r}} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times \mathcal{X}} \|z\|^2 P(dz dy) \right] \\ &= L_2^r(x, \beta), \end{aligned}$$

which concludes the proof of lower semicontinuity of  $L_2^r(x, \beta) = L_2^o(x, \beta)$ .

Next we prove that  $L_2^o(x, \beta)$  is upper semicontinuous. Fix  $(x, \beta) \in \mathbb{R}^{2d}$ . By Lemma 6.6, we know that the optimal control  $\bar{u} = \bar{u}_\beta(x, y)$  exists and can be chosen to be bounded and continuous in  $y$ . Hence, there is a unique invariant measure corresponding to the operator  $\mathcal{L}_{\bar{u}_\beta(x, y), x}^2$  which will be denoted by  $\bar{\mu}_{\bar{u}}(dy|x)$ .

Let  $\{x_n \in \mathbb{R}^d\}$  be such that  $x_n \rightarrow x$  and define a control  $u^n$  by the formula

$$(6.13) \quad \gamma b(x_n, y) + c(x_n, y) + \sigma(x_n, y)u^n(y) = \gamma b(x, y) + c(x, y) + \sigma(x, y)\bar{u}_\beta(x, y).$$

Since  $\sigma(x, y)$  is nondegenerate,  $u^n(y)$  is uniquely defined, continuous in  $y$  and uniformly bounded in  $(n, y)$ , i.e., there exists a constant  $M < \infty$  such that  $\sup_{(n, y) \in \mathbb{N} \times \mathcal{Y}} \|u^n(y)\| \leq M$ . It follows from  $\sigma(x_n, y)[u^n(y) - \bar{u}_\beta(x, y)] = [\sigma(x, y) - \sigma(x_n, y)]\bar{u}_\beta(x, y) + \gamma[b(x, y) - b(x_n, y)] + [c(x, y) - c(x_n, y)]$  that in fact  $u^n(y)$  converges to  $\bar{u}_\beta(x, y)$  uniformly in  $y$ . Since  $u^n(y)$  is bounded and Lipschitz continuous there is a unique invariant measure corresponding to  $\mathcal{L}_{u^n(y), x_n}^2$  which will be denoted by  $\theta^n(dy)$ .

Owing to the definition of  $u^n(y)$  via (6.13), the operator  $\mathcal{L}_{u^n(y), x_n}^2$  takes the form

$$\mathcal{L}_{u^n(y), x_n}^2 = [\gamma b(x, y) + c(x, y) + \sigma(x, y)\bar{u}_\beta(x, y)] \cdot \nabla_y + \gamma \frac{1}{2} \sigma(x_n, y) \sigma(x_n, y)^T : \nabla_y \nabla_y.$$

Hence by Condition 2.2, it follows that  $\theta^n(dy) \rightarrow \bar{\mu}_{\bar{u}}(dy|x)$  in the topology of weak convergence. Let  $\{\beta_n \in \mathbb{R}^d\}$  be defined by

$$\begin{aligned} \beta_n &= \int_{y \in \mathcal{Y}} (\gamma b(x_n, y) + c(x_n, y) + \sigma(x_n, y)u^n(y)) \theta^n(dy) \\ &= \int_{y \in \mathcal{Y}} (\gamma b(x, y) + c(x, y) + \sigma(x, y)\bar{u}_\beta(x, y)) \theta^n(dy). \end{aligned}$$

Then the weak convergence  $\theta^n(dy) \Rightarrow \bar{\mu}_{\bar{u}}(dy|x)$ , the uniform convergence of  $u^n(y)$  to  $\bar{u}_\beta(x, y)$ , and the continuity in  $y$  of the function  $\gamma b(x, y) + c(x, y) + \sigma(x, y)\bar{u}_\beta(x, y)$  imply that  $\beta_n \rightarrow \beta$ . Thus

$$\begin{aligned} \limsup_{n \rightarrow \infty} L_2^o(x_n, \beta_n) &= \limsup_{n \rightarrow \infty} \inf_{(v, \mu) \in \mathcal{A}_{x_n, \beta_n}^{2,r}} \left\{ \frac{1}{2} \int_{\mathcal{Y}} \|v(y)\|^2 \mu(dy) \right\} \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{2} \int_{\mathcal{Y}} \|u^n(y)\|^2 \theta^n(dy) \\ &= \frac{1}{2} \int_{\mathcal{Y}} \|\bar{u}_\beta(x, y)\|^2 \bar{\mu}_{\bar{u}}(dy|x) \\ &= L_2^o(x, \beta). \end{aligned}$$

Line 2 follows from the choice of a particular control. Line 3 follows from the uniform convergence of  $\|u^n(y)\|^2$  to  $\|\bar{u}_\beta(x, y)\|^2$ , the continuity and boundedness of  $\bar{u}_\beta(x, y)$  in  $y$ , and the weak convergence  $\theta^n(dy) \Rightarrow \bar{\mu}_{\bar{u}}(dy|x)$ . Line 4 follows from the fact that  $\bar{u}$  is the control that achieves the infimum in the definition of  $L_2^o(x, \beta)$ .

We have shown that if  $x_n \rightarrow x$  then there exists  $\{\beta_n \in \mathbb{R}^d\}$  such that  $\beta_n \rightarrow \beta$  and  $\limsup_{n \rightarrow \infty} L_2^o(x_n, \beta_n) \leq L_2^o(x, \beta)$ . We claim that in fact the same is true for any sequence  $\beta_n \rightarrow \beta$ . Let  $\delta > 0$  be given. Since  $L_2^o(x, \cdot)$  is finite and convex, we can choose  $\rho^j > 0, \gamma^j \in \mathbb{R}^d, j = 1, \dots, d$ , such that the convex hull of  $\gamma^j \in \mathbb{R}^d, j = 1, \dots, d$  has nonempty interior,  $\sum_{j=1}^d \rho^j = 1, \beta = \sum_{j=1}^d \rho^j \gamma^j$ , and

$$L_2^o(x, \beta) \geq \sum_{j=1}^d \rho^j L_2^o(x, \gamma^j) - \delta.$$

For each  $\gamma^j$  construct a sequence  $\gamma_n^j$  such that  $\gamma_n^j \rightarrow \gamma^j$  and  $\limsup_{n \rightarrow \infty} L_2^o(x_n, \gamma_n^j) \leq L_2^o(x, \gamma^j)$ . Since for all sufficiently large  $n$   $\beta_n$  is in the interior of the convex hull of  $\gamma^j, j = 1, \dots, d$ , there are for all such  $n$   $\rho_n^j > 0$  such that  $\sum_{j=1}^d \rho_n^j = 1, \beta_n = \sum_{j=1}^d \rho_n^j \gamma_n^j$ , and  $\rho_n^j \rightarrow \rho^j$ . By convexity

$$\limsup_{n \rightarrow \infty} L_2^o(x_n, \beta_n) \leq \limsup_{n \rightarrow \infty} \sum_{j=1}^d \rho_n^j L_2^o(x_n, \gamma_n^j) \leq \sum_{j=1}^d \rho^j L_2^o(x, \gamma^j) \leq L_2^o(x, \beta) + \delta.$$

Letting  $\delta \downarrow 0$  concludes the proof of the lemma.  $\square$

**Lemma 6.10.** *The control  $\bar{u} = \bar{u}_\beta(x, y)$  constructed in the proof of Lemma 6.6 is continuous in  $x$ , Lipschitz continuous in  $y$  and measurable in  $(x, y, \beta)$ . Moreover, the invariant measure  $\bar{\mu}_{\bar{u}}(dy|x)$  corresponding to the operator  $\mathcal{L}_{\bar{u}_\beta(x, y), x}^2$  is weakly continuous as a function of  $x$ .*

*Proof.* Recall that

$$\bar{u}_\beta(x, y) = -\sigma(x, y)^T (\nabla_y W_\beta(x, y) - \zeta_\beta(x)),$$

where  $\zeta_\beta(x)$  is a subdifferential of  $L_2^o(x, \beta)$  at  $\beta$ . By Lemma 6.8, the subdifferential of  $L_2^o(x, \beta)$  with respect to  $\beta$  consists only of the gradient  $\nabla_\beta L_2^o(x, \beta)$ . Then continuity of  $\zeta_\beta(x)$  follows from this uniqueness and the joint continuity of  $L_2^o(x, \beta)$  established in Lemma 6.9.

Lipschitz continuity in  $y$  of  $\bar{u}_\beta(x, y)$  was established in Lemma 6.6. We insert  $\bar{u}_\beta(x, y)$  as the optimizer into (6.5). Recall that  $\mathcal{L}_{z, x}^2$  is an operator in  $y$  only and denote by  $\mathcal{L}_{0, x}^2$  the operator  $\mathcal{L}_{z, x}^2$  with the control variable  $z = 0$ . After some rearrangement of terms we get the equation

$$\mathcal{L}_{0, x}^2 \bar{W}_\beta(x, y) - \frac{1}{2} \|\sigma^T(x, y) \nabla_y \bar{W}_\beta(x, y)\|^2 = \bar{H}_\beta(x),$$

where  $\nabla_y \bar{W}_\beta(x, y) = \nabla_y W_\beta(x, y) - \zeta_\beta(x)$  and  $\bar{H}_\beta(x) = \rho(x, \beta) - \zeta_\beta(x) \cdot \beta = \tilde{L}_2^r(x, \zeta_\beta)$ . This is now in the standard form for the Bellman equation of an ergodic control problem. As before a classical sense solution exists, and as a consequence we have the representation

$$\bar{H}_\beta(x) = \inf_v \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \int_0^T \left( \frac{1}{2} \|v_s\|^2 - \zeta_\beta(x) \cdot (\gamma b(x, Y_s) + c(x, Y_s) + \sigma(x, Y_s) v_s) \right) ds,$$

where the infimum is over all progressively measurable controls. Since by Condition 2.2  $b, c$ , and  $\sigma$  are continuous in  $x$  uniformly in  $y$  and since  $\zeta_\beta(x)$  is continuous in  $x$ ,  $\bar{H}_\beta(x)$  is continuous in  $x$ .

A straight forward calculation shows that for any  $x_1, x_2 \in \mathbb{R}^d$ , the function  $\Phi(y) = \bar{W}_\beta(x_1, y) - \bar{W}_\beta(x_2, y)$  satisfies a linear equation. This observation and the general theory for uniformly elliptic equations (see [23]) together with the continuity in  $x$  of  $\bar{H}_\beta(x)$  and  $\zeta_\beta(x)$  and Condition 2.2 imply that  $\nabla_y \bar{W}_\beta(x, y)$  is continuous in  $x$  as well. Hence, due to the continuity of  $\sigma$  we conclude that  $\bar{u}_\beta(x, y) = -\sigma(x, y)^T \nabla_y \bar{W}_\beta(x, y)$  is continuous in  $x$ . Measurability is clear.

Lastly, due to continuity of the optimal control  $\bar{u}$  in  $x$ , Condition 2.2 and uniqueness of  $\bar{\mu}_{\bar{u}(x,y)}(dy|x)$  for each  $x$ , we conclude that  $\bar{\mu}_{\bar{u}(x,y)}(dy|x)$  is weakly continuous as a function of  $x$  (see, e.g., Section 3 in [3]).  $\square$

## 7. LAPLACE PRINCIPLE UPPER BOUND FOR REGIME 3

In this section we discuss the Laplace principle upper bound for Regime 3. For notational convenience we drop the superscript 3 from  $\bar{X}^3$  and  $P^3$ .

We consider the general multidimensional case when  $c(x, y) = c(y)$  and  $\sigma(x, y) = \sigma(y)$ . In Remark 7.1 we discuss the case when the functions  $c$  and  $\sigma$  depend on  $x$  as well. In Subsection 7.1 we consider the  $d = 1$  case. For  $d = 1$  we can establish the LDP when the coefficients depend on  $x$  as well and we provide an alternative expression for the rate function together with a control that nearly achieves the large deviations lower bound at the prelimit level. An easy computation shows that this alternate expression is equivalent to the corresponding expression in [20] for  $b(x, y) = b(y)$ ,  $c(x, y) = c(y)$  and  $\sigma(x, y) = \sigma(y)$  for  $d = 1$ .

*Remarks on the proof of Laplace principle upper bound for Regime 3.* For each  $\epsilon > 0$ , let  $X^\epsilon$  be the unique strong solution to (1.1). To prove the Laplace principle upper bound we must show that for all bounded, continuous functions  $h$  mapping  $\mathcal{C}([0, 1]; \mathbb{R}^d)$  into  $\mathbb{R}$

$$\limsup_{\epsilon \downarrow 0} -\epsilon \ln \mathbb{E}_{x_0} \left[ \exp \left\{ -\frac{h(X^\epsilon)}{\epsilon} \right\} \right] \leq \inf_{(\phi, P) \in \mathcal{V}} \left[ \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P(dz dy dt) + h(\phi) \right].$$

Define

$$I(P) = \frac{1}{2} \int_{\mathcal{Z} \times \mathcal{Y} \times [0, 1]} \|z\|^2 P(dz dy dt).$$

Let  $\eta > 0$  be given and consider  $(\psi, \bar{P}) \in \mathcal{V}$  with  $\psi_0 = x_0$  such that

$$I(\bar{P}) + h(\psi) \leq \inf_{(\phi, P) \in \mathcal{V}} [I(P) + h(\phi)] + \eta < \infty.$$

We claim that there is a family of controls  $\{\bar{u}^\epsilon, \epsilon > 0\}$  such that

$$(\bar{X}^\epsilon, \bar{P}^\epsilon, \Delta) \xrightarrow{\mathcal{D}} (\bar{X}, \bar{P}) \text{ and } \bar{X} = \psi \text{ w.p.1,}$$

where  $(\bar{X}^\epsilon, \bar{P}^\epsilon, \Delta)$  is constructed using  $\bar{u}^\epsilon$ . With this at hand the result easily follows.

The claim follows from the results in Section 3 in [21] and Section 4 in [13]. Note that in the case considered here, the fast motion is restricted to remain in a compact set at all times, the dynamics are affine in the control,  $\sigma$  is uniformly nondegenerate and the functions  $c$  and  $\sigma$  do not depend on  $x$ . For the construction of the control and precise statements we refer the reader to [21, 13].  $\square$

**Remark 7.1.** (i) *The difficulties that arise in Regime 3 are due to the fact that one has to average with respect to a first order operator. In this case uniqueness of an invariant measure is not guaranteed and is actually difficult to verify in practice.*

(ii) *Suppose that the functions  $c$  and  $\sigma$  depend on  $x$  as well. It turns out that under some additional Lipschitz type conditions in  $x$ , one can still use the methodology in [21, 13]. These conditions are automatically satisfied for any admissible control if the functions  $c(x, y)$  and  $\sigma(x, y)$  do not depend on  $x$ . However, we were unable to verify them when the coefficients depend on  $x$  without imposing any further restrictions on the class of controls under consideration. For a more detailed discussion see [21, 13].*



**7.1. Regime 3: An alternative expression for the rate function in dimension  $d = 1$ .** In this subsection we give an alternative expression of the rate function for Regime 3 in dimension  $d = 1$ . The proof is analogous to the proof of the statement for Regime 2. We therefore only state the result without proving it. The reason one can prove the LDP for  $d = 1$  with the coefficients depending on  $x$  is that the invariant measure takes an explicit form. Then, the local rate function is the value function to a calculus of variations problem which can be analyzed by standard techniques. In particular, because everything can be written explicitly, we can easily prove that the infimum of this variational problem is attained at a control  $\bar{u}$  for which the corresponding ODE has a unique invariant measure.

Consider a control  $v(\cdot) : \mathcal{Y} \mapsto \mathbb{R}$ . Without loss of generality one can restrict attention to controls that give nonzero velocity everywhere. The control  $v$  might depend on  $(t, x)$  as well, but we omit writing it for notational convenience. Decomposing the limiting occupation measure as stochastic kernels (as it was done for Regimes 1 and 2) and fixing the velocity  $\beta = \dot{\phi}_t$ , equations (2.4) and (2.5) with  $(\lambda, \mathcal{L}_{z,x}) = (\lambda_3, \mathcal{L}_{z,x}^3)$  imply that the corresponding invariant measure  $\mu_v(dy)$  that satisfies (2.5) takes the form

$$\mu_v(dy) = \frac{\beta}{c(x, y) + \sigma(x, y)v(y)} dy.$$

For  $x, \beta \in \mathbb{R}$  define

$$J_{x,\beta}(v) = \frac{1}{2} \int_{\mathbb{T}} |v(y)|^2 \frac{\beta}{c(x, y) + \sigma(x, y)v(y)} dy,$$

and the local rate function

$$L_3^o(x, \beta) = \inf_v \left\{ J_{x,\beta}(v) : \int_{\mathbb{T}} \frac{\beta}{c(x, y) + \sigma(x, y)v(y)} dy = 1 \right\}.$$

**Theorem 7.2.** *Assume Condition 2.2 and that we are considering Regime 3. Let  $\{X^\epsilon, \epsilon > 0\}$  be the 1-dimensional diffusion process that satisfies (1.1). Then  $\{X^\epsilon, \epsilon > 0\}$  satisfies the large deviations principle with rate function*

$$S(\phi) = \begin{cases} \int_0^1 L_3^o(\phi_s, \dot{\phi}_s) ds & \text{if } \phi \in \mathcal{C}([0, 1]; \mathbb{R}) \text{ is absolutely continuous} \\ +\infty & \text{otherwise.} \end{cases}$$

We conclude this section with the following corollary. As can be easily seen from the form of  $L_3^o$  in Theorem 7.2, in the case  $c(x, y) = 0$  one obtains a closed form expression for the rate function.

**Corollary 7.3.** *In addition to the conditions of Theorem 7.2, assume that  $c(x, y) = 0$ . Then  $\{X^\epsilon, \epsilon > 0\}$  satisfies the large deviations principle with rate function*

$$S(\phi) = \begin{cases} \frac{1}{2} \int_0^1 |\dot{\phi}_s|^2 \left| \int_{\mathbb{T}} (\sigma^2(\phi_s, y))^{-1/2} dy \right|^2 ds & \text{if } \phi \in \mathcal{C}([0, 1]; \mathbb{R}) \text{ is absolutely continuous,} \\ +\infty & \text{otherwise.} \end{cases}$$

## 8. ACKNOWLEDGEMENTS

We would like to thank Hui Wang for his initial involvement in this project.

## REFERENCES

- [1] O. Alvarez, M. Bardi, Viscosity solutions methods for singular perturbations in deterministic and stochastic control, *SIAM Journal on Control and Optimization* Vol. 40, Issue 4, (2001), pp. 1159-1188.
- [2] M. Arisawa, P.-L. Lions, On ergodic stochastic control, *Communications in Partial Differential Equations*, Vol. 23, Issue 11, (1998), pp. 333-358.
- [3] Z. Artstein, A. Vigodner, Singularly perturbed ordinary differential equations with dynamic limits, *Proceedings of the Royal Society of Edinburgh*, Section A, Vol. 126, (1996), pp. 541-569.

- [4] Z. Artstein, V. Gaitsgory, Tracking fast trajectories along a slow dynamics: A singular perturbation analysis, *SIAM Journal on Control and Optimization*, Vol. 35, No. 5, (1997), pp. 1487-1507.
- [5] Z. Artstein, V. Gaitsgory, The value function of singularly perturbed control systems, *Applied Mathematics and Optimization*, Vol. 41, Issue 3, (2000), pp. 425-445.
- [6] A. Ansari, Mean first passage time solution of the Smoluchowski equation: Application of relaxation dynamics in myoglobin, *Journal of Chemical Physics*, Vol. 112, No. 5, (2000), pp. 2516-2522.
- [7] A. Arapostathis, V. S. Borkar, M. K. Ghosh, *Ergodic Control of Diffusion Processes*, Cambridge University, press for publication.
- [8] P. Baldi, Large deviations for diffusions processes with homogenization and applications, *Annals of Probability*, Vol. 19, No. 2, (1991), pp. 509-524.
- [9] A. Bensoussan, J.L. Lions, G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, Vol 5, Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam, 1978.
- [10] A. Bensoussan, J. Frehse, On Bellman equations of ergodic control in  $\mathbb{R}^n$ , *Journal für die Reine und Angewandte Mathematik*, Vol. 429, (1992), pp. 125-160.
- [11] V. Borkar, *Optimal Control of Diffusion Processes*, Pitman research notes in mathematics series 203, Longman scientific and technical, Harlow, UK, 1989.
- [12] V. Borkar, V. Gaitsgory, On existence of limit occupational measures set of a controlled stochastic differential equation, *SIAM Journal on Control and Optimization*, Vol 44, No. 4, (2005), pp. 1436-1473.
- [13] V. Borkar, V. Gaitsgory, Averaging of singularly perturbed controlled stochastic differential equations, *Applied Mathematics and Optimization*, Vol 56, No. 2, (2007), pp. 169-209.
- [14] M. Boué, P. Dupuis, A variational representation for certain functionals of Brownian motion, *Annals of Probability*, Vol. 26, Issue 4, (1998), pp. 1641-1659.
- [15] M. Boué, P. Dupuis, R. S. Ellis, Large deviations for small noise diffusions with discontinuous statistics, *Probability Theory and Related Fields*, Vol. 116, Issue 1, (2000), pp. 125-149.
- [16] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways and the energy landscape of protein folding: A synthesis, *Proteins*, Vol. 21, Issue 3, (1995), pp. 167-195.
- [17] R. Buckdahn, Y. Hu, S. Peng, Probabilistic approach to homogenization of viscosity solutions of parabolic PDE's, *Nonlinear Differential Equations and Applications*, Vol. 6, No. 4, (1999), pp. 395-411.
- [18] P. Dupuis, R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, New York, 1997.
- [19] S.N. Eithier, T.G. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York, 1986.
- [20] M. Freidlin, R. Sowers, A comparison of homogenization and large deviations, with applications to wavefront propagation, *Stochastic Process and Their Applications*, Vol. 82, Issue 1, (1999), pp. 23-52.
- [21] V. Gaitsgory, On a representation of the limit occupational measures set of a control system with applications to singularly perturbed control systems, *SIAM Journal on Control and Optimization*, Vol. 43, No. 1, (2004), pp. 325-340.
- [22] V. Gaitsgory, M.-T. Nguyen, Multiscale singularly perturbed control systems: Limit occupational measures sets and averaging, *SIAM Journal on Control and Optimization*, Vol. 41, No. 3, (2002), pp. 954-974.
- [23] D. Gilbarg, N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd Ed., Springer-Verlag, (1983).
- [24] C. Hyeon, D. Thirumalai, Can energy landscapes roughness of proteins and RNA be measured by using mechanical unfolding experiments?, *Proc. Natl. Acad. Sci. USA*, Vol. 100, No. 18, (2003), pp. 10249-10253.
- [25] K. Horie, H. Ishii: Simultaneous effects of homogenization and vanishing viscosity in fully nonlinear elliptic equations, *Funkcialaj Ekvacioj*, Vol. 46, No. 1, (2003), pp. 63-88.
- [26] H. Kaise, S.J. Sheu, On the structure of solutions of ergodic type Bellman equation related to risk-sensitive control, *Annals of Probability*, Vol. 34, No. 1, (2006), pp. 284-320.
- [27] R.Z. Khasminskii, Ergodic properties of recurrent diffusion processes and stabilization of the solution to the Cauchy problem for parabolic equations, *Theory of Probability and its Applications*, Vol. 5, Issue 2, (1960), pp. 179-196.
- [28] R.Z. Khasminskii, *Stochastic Stability of Differential Equations*, Kluwer Academic Publishers, 2nd Ed., 1980.
- [29] T.G. Kurtz, R.H. Stockbridge, Existence of Markov controls and characterization of optimal Markov Controls, *SIAM Journal on Control and Optimization*, Vol 36, No 2, (1998), pp. 609-653.
- [30] H. J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser, Boston-Basel-Berlin, (1990).
- [31] H. J. Kushner, Large deviations for two-time-scale diffusions with delays, submitted, (2009).
- [32] H. J. Kushner, P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, 2nd Ed., 2000.

- [33] R. Lipster, Large deviations for two scaled diffusions, *Probability Theory and Related Fields*, Vol. 106, No. 1, (1996), pp. 71-104.
- [34] S. Lifson, J.L. Jackson, On the self-diffusion of ions in a polyelectrolyte solution, *Journal of Chemical Physics*, Vol. 36, (1962), pp. 2410-2414.
- [35] D. Mondal, P.K. Ghosh, D.S. Ray, Noise-induced transport in a rough ratchet potential, *Journal of Chemical Physics*, Vol. 130, (2009), pp. 074703.1-074703.7.
- [36] E. Pardoux, A.Yu. Veretennikov, On Poisson equation and diffusion approximation 2, *Annals of Probability*, Vol. 31, No. 3, (2003), pp. 1166-1192.
- [37] M. Robin, Long-time average cost control problems for continuous time Markov processes: A survey, *Acta Applicandae Mathematicae*, 1, (1983), pp. 281-299.
- [38] R.T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, (1970).
- [39] J.G. Saven, J. Wang, P.G. Wolynes, Kinetics of protein folding: The dynamics of globally connected rough energy landscapes with biases, *Journal of Chemical Physics*, Vol. 101, No. 12, (1994), pp. 11037-11043.
- [40] A. Yu. Veretennikov, On large deviations in the averaging principle for SDEs with a “full dependence”, correction, [arXiv:math/0502098v1](#) [[math.PR](#)] (2005). Initial article in *Annals of Probability*, Vol. 27, No. 1, (1999), pp. 284-296.
- [41] A. Yu. Veretennikov, On large deviations for SDEs with small diffusion and averaging, *Stochastic Processes and their Applications*, Vol. 89, Issue 1, (2000), pp. 69-79.
- [42] R. Zwanzig, Diffusion in a rough potential, *Proc. Natl. Acad. Sci. USA*, Vol. 85, (1988), pp. 2029-2030.